



Univerzitet u Novom Sadu  
Ekonomski fakultet u Subotici  
Poslovna informatika

# **Sentiment u sadržajima sa društvenih mreža kao instrument unapređenja poslovanja visokoškolskih institucija**

Doktorska disertacija

Mentor: **Prof. dr Saša Bošnjak**

Kandidat: **Olivera Grljević**

Subotica, 2016. godine

## KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj, RBR:	
Identifikacioni broj, IBR:	
Tip dokumentacije, TD:	Monografska dokumentacija
Tip zapisa, TZ:	Tekstualni štampani materijal
Vrsta rada, VR:	Doktorska disertacija
Ime i prezime autora, AU:	Olivera Grljević
Mentor, MN:	dr Saša Bošnjak, redovni profesor
Naslov rada, NR:	Sentiment u sadržajima sa društvenih mreža kao instrument unapređenja poslovanja visokoškolskih institucija
Jezik publikacije, JP:	Srpski
Jezik izvoda, JI:	Srpski/engleski
Zemlja publikovanja, ZP:	Srbija
Uže geografsko područje, UGP:	Vojvodina
Godina, GO:	2016.
Izdavač, IZ:	Autorski reprint
Mesto i adresa, MA:	Subotica, Segedinski put 9-11
Fizički opis rada, FO: (broj poglavlja / stranica / slika / grafikona / tabela / referenci / priloga)	(5 / 257 / 55 / 39 / 71 / 185 / 12)
Naučna oblast, NO:	Ekonomske nauke – Poslovna informatika
Naučna disciplina, ND:	Rudarenje podataka ( <i>data mining</i> )
Predmetna odrednica, ključne reči, PO:	društveni mediji, društvene mreže, onlajn recenzije, anotacija, korpus, rudarenje podataka, sentiment analiza, visoko obrazovanje
UDK	
Čuva se, ČU:	Biblioteka Ekonomskog fakulteta u Subotici, Segedinski put 9-11, 24000 Subotica

Važna napomena, VN:	/
Izvod, IZ:	<p>Istraživanje, koje je predmet doktorske disertacije, daje sveobuhvatni pristup modelovanju i automatizaciji analize sentimenta sadržanog u studentskim recenzijama nastavnog osoblja na sajtovima društvenih mreža i medija.</p> <p>Recenzije profesora, koji predaju na različitim visokoškolskim institucijama u Srbiji, preuzete su sa sajta <i>Oceni profesora</i>. Kao rezultat istraživanja, primenom postupka manuelne anotacije izgrađen je korpus na srpskom jeziku za domen visokog obrazovanja. Recenzije su ručno obogaćene informacijama o aspektu recenziranja, sentiment orijentaciji, intenzitetu iskazanog sentimenta, sentiment rečima koje utiču na orijentaciju sveukupno iskazanog sentimenta, negaciji i opsegu njenog važenja.</p> <p>Ovako pripremljen i anotiran korpus upotrebljen je u sentiment analizi za obučavanje algoritama nadgledanog učenja (Naive Bayes, Support Vector Machines i k-Nearest Neighbor). Na osnovu anotiranih sentiment reči, ključnih reči negacije, kao i na osnovu zapažanja izvedenih tokom anotacije kreirano je pet rečnika primenjenih u sentiment analizi zasnovanoj na leksikonima. U oba slučaja zadatak klasifikacije sentimenta realizovan je na dva nivoa granularnosti: na nivou dokumenta i na nivou rečenice.</p> <p>U analizi mogućnosti upotrebe razvijenih modela u poslovanju visokoškolskih institucija, na primeru Ekonomskog fakulteta u Subotici, izvršena je automatizovana klasifikacija sentimenta studenata na pozitivan, negativan i neutralan, analizirani su različiti aspekti poslovanja prema sentiment polaritetu, primenom asocijativnih pravila identifikovana je međupovezanost najčešće korišćenih pozitivnih i negativnih pojmova, kao i analiza trenda u sentimentu kroz vreme. Realizovane analize omogućile su poređenje onlajn reputacije institucije sa onlajn reputacijom konkurencije, identifikovanje aspekta koji predstavlja usko grlo u poslovanju visokoškolske institucije, identifikovanje izvora zadovoljstva i nezadovoljstva studenata, kao i praćenje oscilacija u studentskom sentimentu tokom</p>

	vremena.
Datum prihvatanja teme od strane Senata, DP:	12.2.2015. godine
Datum odbrane, DO:	
Članovi komisije, KO: (ime i prezime / titula / zvanje / naziv organizacije / status)	predsednik:  član:  član:

## KEY WORD DOCUMENTATION

Accession number, ANO:	
Identification number, INO:	
Document type, DT:	Monograph documentation
Type of record, TR:	Textual printed material
Contents code, CC:	PhD thesis
Author, AU:	Olivera Grljević
Mentor, MN:	Saša Bošnjak, PhD Full Professor
Title, TI:	Sentiment in social networks as means of business improvement of higher education institutions
Language of text, LT:	Serbian
Language of abstract, LA:	Serbian/English
Country of publication, CP:	Serbia
Locality of publication, LP:	Vojvodina
Publication year, PY:	2016
Publisher, PU:	Autor's reprint
Publication place, PP:	9-11 Segedinski put Street, 24000 Subotica
Physical description, PD: (number of chapters / pages / illustrations / graphs / tables / references / appendices)	(5 / 257 / 55 / 39 / 71 / 185 / 12)
Scientific field, SF:	Economics – Business Informatics
Scientific discipline, SD:	Data mining
Subject, Key words, SKW:	social media, social networks, online reviews, annotation, corpus, data mining, sentiment analysis, higher education
UC	
Holding data, HD:	Faculty Library, 9-11 Segedinski put Street, 24000 Subotica

Note, N:	/
Abstract, AB:	<p>Doctoral dissertation provides a comprehensive approach to modelling and automation of analysis of sentiments contained in student reviews of teaching staff available on social media and social networking sites.</p> <p>Reviews of professors, who teach at various institutions of higher education in Serbia, have been collected from the site <i>Oceni profesora</i>. The process of manual annotation was implemented to develop corpus in Serbian language for the domain of higher education. Reviews of professors were manually enriched with information on aspect, sentiment orientation, the intensity of the expressed sentiment, sentiment words that affect orientation of the overall sentiment, negation, and the extent of its validity.</p> <p>Thus prepared and annotated corpus was used in sentiment analysis for training the supervised algorithms (Naïve Bayes, Support Vector Machines and K-Nearest Neighbor). Five dictionaries were developed based on the annotation of sentiment words, negation keywords, and observations derived during the annotation process, which were consequently used in sentiment analysis based on lexicons. In both cases, sentiment classification was carried out at two levels of granularity: the document level and sentence level.</p> <p>Analysis of the utilization possibilities of the developed models has been conducted for the Faculty of Economics in Subotica. We conducted automated classification of students' sentiment, analysis of various business aspects according to sentiment polarity, association rules which revealed relations between the most frequently used positive and negative terms, as well as the analysis of trends in sentiment over time. These analysis enabled comparison of online reputation among competitors, identification of aspects that represent a bottleneck of business operations of higher education institutions, identification of satisfaction and dissatisfaction sources among students, as well as monitoring of fluctuations in student sentiment over time.</p>
Accepted on Senate on, AS:	12 <sup>th</sup> February 2015.
Defended, DE:	

Thesis Defend Board, DB:	president:  member:  member:
--------------------------	------------------------------------------

## Apstrakt

Razvoj veb 2.0 je omogućio svim korisnicima interneta da postanu aktivni kreatori sadržaja na vebu i da učestvuju u njegovom oblikovanju. Ova pojava je doprinela i velikoj ekspanziji društvenih mreža i medija kao instrumenta onlajn komuniciranja. Korisnički generisani sadržaji, dostupni na sajtovima društvenih mreža i medija, imaju snažan uticaj na oblikovanje mišljenja potrošača – pojedinci ih konsultuju s ciljem donošenja informisane odluke o kupovini. Otuda, negativni komentari o kompaniji, brendu ili konkretnom proizvodu/usluzi mogu imati drastične efekte na poslovanje kompanije.

Sentiment analiza predstavlja istraživačko područje koje kompanijama omogućava analizu mišljenja javnosti, sentimenta, stavova i emocija usmerenih ka određenom entitetu. Na ovaj način kompanije upoznaju bazu svojih potrošača na višem nivou, koji uključuje i emocije, što nadalje omogućava optimizaciju marketing poruke, predviđanje trendova, razvoj proizvoda/usluga i praćenje onlajn reputacije.

Analogno drugim oblastima poslovanja, i u oblasti visokog obrazovanja potrebno je derivirati saznanja iz velikih količina podataka i informacija, koje se razmenjuju na sajtovima društvenih mreža i medija. Saznanja o mišljenju, stavovima i osećanjima studenata u vezi sa institucijom, nastavnim sadržajima i predmetima, ophođenjem nastavnog i nenastavnog osoblja i drugim aspektima školovanja pružaju visokoškolskim institucijama osnovu za unapređenje poslovanja, komunikacije i odnosa sa studentima. Da bi se sentiment analiza uspešno sproveda, neophodno je stvoriti adekvatne uslove za njenu realizaciju, koji su obezbeđeni realizovanim istraživanjem.

Istraživanje, koje je predmet doktorske disertacije, daje sveobuhvatni pristup modelovanju i automatizaciji analize sentimenta sadržanog u studentskim recenzijama nastavnog osoblja na sajtovima društvenih mreža i medija. Budući da skupovi podataka, tzv. korpusi, neophodni za realizaciju sentiment analize sadržaja na srpskom jeziku ne postoje ili nisu javno dostupni, primenom postupka manuelne anotacije izgrađen je korpus na srpskom jeziku za domen visokog obrazovanja, što predstavlja jedan od rezultata istraživanja. Relevantni podaci su prikupljeni sa sajta *Oceni profesora* – 3863 recenzija profesora, koji predaju na različitim visokoškolskim institucijama u Srbiji, sa 6896 rečenica. Anotaciji korpusa je prethodio korak preprocesiranja preuzetog sadržaja, koji podrazumeva uklanjanje sadržaja pisanog na stranom jeziku, korigovanje pravopisnih grešaka i dekompoziciju recenzija na rečenice. Usled ograničenja koja nameće srpski jezik, postojeći alati za razdvajanje sadržaja na rečenice nisu se mogli upotrebiti, te je razvijen razdelnik rečenica prilagođen potrebama našeg jezika, koji ima mogućnost i daljeg proširenja.

Anotaciju preprocesiranog sadržaja realizovala su četiri anotatora prateći definisanu shemu anotacije i uputstvo za anotiranje, koji su pripremljeni tokom izrade disertacije. Prikupljene recenzije nastavnog osoblja ručno su obogaćene informacijama o aspektu recenziranja (profesor,



odnos prema studentima, predavanja, predmet, materijali, organizacija predmeta i drugi aspekt), sentiment orijentaciji (pozitivna, negativna i neutralna), intenzitetu iskazanog sentimenta, sentiment rečima koje utiču na orijentaciju sveukupno iskazanog sentimenta, negaciji i opsegu njenog važenja.

Pripremljen i anotiran korpus upotrebljen je u analizama, čiji je fokus utvrđivanje i klasifikovanje sentimenta. Klasifikacija sentimenta je realizovana na dva nivoa granularnosti: na nivou dokumenta i na nivou rečenice. Za različite nivoe analize eksperimentisano je pristupima zasnovanim na leksikonima i pristupima nadgledanog učenja. Algoritmi nadgledanog učenja (Naïve Bayes, Support Vector Machines i k-Nearest Neighbor) obučavani su za automatizovanu klasifikaciju pozitivnih i negativnih recenzija. Na osnovu anotiranih sentiment reči, ključnih reči negacije, kao i na osnovu zapažanja izvedenih tokom anotacije kreirano je pet rečnika sa kojima je eksperimentisano u sentiment analizi zasnovanoj na leksikonima. U doktorskoj disertaciji data je detaljna evaluacija pristupa i diskusija rezultata. Razvijeni modeli stvorili su uslove za analizu onlajn reputacije visokoškolske institucije i za identifikovanje preferencija i izvora nezadovoljstva studenata.

U analizi mogućnosti upotrebe razvijenih modela u poslovanju visokoškolskih institucija, na primeru Ekonomskog fakulteta u Subotici, izvršena je automatizovana klasifikacija sentimenta studenata na pozitivan, negativan i neutralan, analizirani su različiti aspekti poslovanja prema sentiment polaritetu, primenom asocijativnih pravila identifikovana je međupovezanost najčešće korišćenih pozitivnih i negativnih pojmova, kao i analiza trenda u sentimentu kroz vreme. Realizovane analize omogućile su poređenje onlajn reputacije institucije sa onlajn reputacijom konkurencije, identifikovanje aspekta koji predstavlja usko grlo u poslovanju visokoškolske institucije, identifikovanje izvora zadovoljstva i nezadovoljstva studenata, kao i praćenje oscilacija u studentskom sentimentu tokom vremena. Rezultati istraživanja, prikazani u doktorskoj disertaciji, jasno ukazuju na mogućnosti za poboljšanje poslovanja visokoškolske institucije kroz primenu saznanja stečenih sentiment analizom relevantnih tekstualnih sadržaja, pisanih na srpskom jeziku.

## Abstract

The development of web 2.0 enabled all Internet users to become active web content creators as well as to participate in its design. This phenomenon has contributed the expansion of social networks and social media as online communication instruments. User-generated contents available on social networks and social media sites have powerful influence on the consumer's opinion - individuals consult them to make more informed purchase decisions. Therefore the negative comments regarding the company, brand or a specific product/service can have drastic effects on the company's business operations.

Sentiment analysis is a research area that enables companies to analyze public opinion, sentiment, attitudes and emotions directed towards a particular entity, and to familiarize with its customer base at a higher level which includes the emotions. This enables optimization of marketing messages, trend forecasting, development of products/services and online reputation monitoring.

In the field of higher education, similarly to other areas of business, it is necessary to derive insights from large amounts of data and information shared on social networks and media. Insights regarding students' views, attitudes and feelings related to the institution, subject, the relationship of teaching and non-teaching staff to students, or other aspects of business, provide the basis for improving business of higher education institution, communication and relations with students. In order to successfully implement sentiment analysis it is necessary to create adequate conditions for its implementation, which are provided through implemented research.

Doctoral dissertation provides a comprehensive approach to modelling and automation of sentiment analysis contained in student reviews of teaching staff. Datasets, i.e. corpora, necessary for realization of the sentiment analysis of content written in the Serbian language, do not exist or are not publicly available. The process of manual annotation was implemented to develop corpus in Serbian language for the domain of higher education, which represents one of the research results. Relevant data has been collected from the site *Oceni profesora* - 3863 reviews of professors, who teach at various institutions of higher education in Serbia, with the 6896 sentences. Preprocessing of the retrieved content precedes the annotation. This involves the removal of content written in a foreign language, correcting spelling mistakes and sentence splitting in reviews. Available tools for sentence splitting could not be used due to restrictions imposed by the Serbian language. Therefore sentence splitter tailored to the needs of our language has been developed, which also has the possibility of further expansion.

The annotation of the preprocessed content was carried out by four annotators that followed defined annotation schema and annotation guidelines, which were prepared in the early stages of the research. Retrieved reviews of professors were manually enriched with information on reviewed aspects (professor, relations towards students, lectures, subject, study materials,

organization of subjects, and other aspect), sentiment orientation (positive, negative and neutral), the intensity of the expressed sentiment, sentiment words that affect orientation of the overall sentiment, negation, and the extent of its validity.

Thus prepared and annotated corpus was used in analysis with the focus of identification and classification of sentiment. Sentiment classification was carried out at two levels of granularity: the document level and sentence level. For these different levels of analysis, we experiment with dictionary-based and supervised approaches. Supervised learning algorithms (Naïve Bayes, Support Vector Machines and K-Nearest Neighbor) were trained for automated classification of positive and negative reviews. Five dictionaries were developed based on the annotation of sentiment words, negation keywords, and observations derived during the annotation process, which were consequently used in experiments in sentiment analysis based on lexicons. The thesis provides detailed evaluation of deployed approaches and discussion of the results. The developed models have created conditions for the analysis of online reputation of higher education institutions and facilitate the identification of preferences and dissatisfaction of students.

Analysis of the utilization possibilities of the developed models has been conducted for the Faculty of Economics in Subotica. We conducted automated classification of students` sentiment, analysis of various business aspects according to sentiment polarity, association rules which revealed relations between the most frequently used positive and negative terms, as well as the analysis of trends in sentiment over time. These analysis enabled comparison of online reputation among competitors, identification of aspects that represent a bottleneck of business operations of higher education institutions, identification of satisfaction and dissatisfaction sources among students, as well as monitoring of fluctuations in student sentiment over time. The findings, presented in a doctoral thesis, clearly suggest possibilities for improving the business of higher education institutions through the application of knowledge acquired through the sentiment analysis of relevant text written in Serbian language.

# SADRŽAJ

<b>1. UVOD</b> .....	<b>1</b>
1.1. Društvene mreže i društveni mediji – pojam, mesto i uloga u savremenom društvu .....	2
1.2. Analiza podataka o mišljenjima i sentiment .....	7
1.3. Opis istraživanja i pristupa .....	10
1.4. Predmetna terminologija .....	14
<b>2. SENTIMENT ANALIZA</b> .....	<b>19</b>
2.1. Sentiment analiza i mišljenja o proizvodima i uslugama .....	20
2.2. Sentiment analiza u finansijskom sektoru .....	25
2.3. Primena sentiment analize u poslovnoj inteligenciji.....	27
2.4. Ostale primene sentiment analize.....	28
2.5. Sentiment analiza u visokom obrazovanju.....	30
<b>3. IMPLEMENTACIJA SENTIMENT ANALIZE</b> .....	<b>33</b>
3.1. Metodologija sentiment analize .....	34
3.2. Izazovi u sentiment analizi .....	36
3.2.1. Domenska i jezička zavisnost.....	36
3.2.2. Ekstrakcija atributa .....	37
3.2.3. Detekcija i rukovanje negacijom, sarkazmom i ironijom .....	39
3.2.4. Utvrđivanje polariteta .....	40
3.2.5. Utvrđivanje subjektivnosti dokumenta .....	40
3.3. Metode, tehnike i pristupi u sentiment analizi .....	41
3.3.1. Algoritmi nadgledanog učenja u sentiment analizi .....	42
3.3.1.1. Stabla odlučivanja.....	42
3.3.1.2. Linearni klasifikatori: Neuronske mreže i Mašine potpornog vektora .....	43
3.3.1.3. Klasifikator zasnovan na pravilima .....	44
3.3.1.4. Probabilistički klasifikatori.....	44
3.3.1.5. Klasifikator k-najbližih suseda .....	46
3.3.2. Primena nenadgledanog učenja u sentiment analizi .....	46
3.3.3. Klasifikacija zasnovana na leksikonima.....	48
3.3.4. Sentiment analiza zasnovana na aspektima .....	49
3.3.4.1. Sumiranje mišljenja .....	51
<b>4. PRIMENA SENTIMENT ANALIZE U VISOKOM OBRAZOVANJU</b> .....	<b>53</b>
4.1. Metodološki koraci istraživanja .....	56
4.1.1. Prikupljanje i preprocesiranje podataka za osnovni skup iz društvenih medija .....	58
4.1.2. Modelovanje sadržaja za potrebe anotacije korpusa .....	66

4.1.3. Kreiranje anotiranog korpusa za sentiment analizu pominjanja visokoškolskih institucija.....	69
4.1.3.1. Proces anotacije korpusa.....	72
4.1.3.2. Procena stepena saglasnosti anotatora .....	77
4.1.3.3. Statistička eksploracija i prikaz anotiranog korpusa.....	85
4.1.4. Formiranje rečnika sentiment pojmova .....	97
4.1.5. Otkrivanje preovlađujućeg sentimenta primenom inteligentnih metoda i tehnika .....	99
4.1.5.1. Algoritmi mašinskog učenja korišćeni za otkrivanje sentimenta .....	99
4.1.5.2. Sentiment analiza zasnovana na leksikonima .....	103
4.1.6. Interpretacija i prikaz rezultata .....	106
4.1.6.1. Evaluacija klasifikatora izgrađenih algoritmima mašinskog učenja.....	109
4.1.6.2. Evaluacija modela sentiment analize zasnovane na leksikonima .....	116
4.1.6.3. Diskusija rezultata.....	129
4.2. Mogućnost unapređenja poslovanja visokoškolskih institucija primenom zaključaka analize.....	132
<b>5. ZAKLJUČAK .....</b>	<b>147</b>
<b>LITERATURA .....</b>	<b>158</b>
<b>PRILOZI.....</b>	<b>174</b>
A. Preuzimanje sadržaja sa sajta Oceni profesora.....	174
B. Razdelnik rečenica.....	175
C. XML shema anotacije.....	176
D. Uputstvo za anotaciju korpusa .....	177
D.1. Atribut Aspekt.....	179
D.2. Atribut Sentiment polaritet .....	186
D.3. Atribut Intenzitet sentimenta .....	189
D.4. Atribut Sentiment reč .....	191
D.5. Anotiranje negacije .....	192
D.6. Alat za anotaciju .....	193
D.7. Postupak anotacije .....	196
D.8. Čuvanje anotacije i nastavak rada.....	202
E. Procena stepena saglasnosti anotatora .....	204
F. Sentiment reči identifikovane u korpusu .....	209
F.1. Jedinstveni oblici pozitivnih sentiment reči u skupu za obučavanje .....	209
F.2. Jedinstveni oblici negativnih reči u skupu za obučavanje .....	217
F.3. Jedinstveni oblici pozitivnih sentiment reči u skupu za testiranje .....	227
F.4. Jedinstveni oblici negativnih sentiment reči u skupu za testiranje .....	232
G. Intenzifikatori sentimenta identifikovani u korpusu .....	238

H. Domensko specifične fraze identifikovane u korpusu .....	239
I. Ključne reči negacije identifikovane u korpusu.....	240
J. Modeli klasifikacije sentimenta zasnovani na algoritmima nadgledanog učenja .....	241
J.1. Naïve Bayes klasifikator .....	241
J.2. Support vector machines klasifikator .....	244
J.3. k-Nearest Neighbor klasifikator .....	246
K. Sentiment analiza na osnovu leksikona: Primenjeni algoritmi .....	248
L. Evaluacione mere binarne klasifikacije sentimenta zasnovane na leksikonima .....	250

# 1. UVOD

Razvoj veb 2.0 (engl. *Web 2.0*) tehnologija i ekspanzija upotrebe sajtova društvenih mreža i medija izmenili su profil tipičnog potrošača i omogućili kompanijama i svakom pojedincu sa internet konekcijom pristup velikim količinama informacija. Pored zabavnog karaktera, sajtovi društvenih medija i mreža predstavljaju platformu za slobodno iskazivanje mišljenja pojedinaca i za razmenu potrošačkih iskustava, te dobijaju ulogu savremenog marketinga „od usta do usta“. Pri donošenju odluke o kupovini proizvoda i usluga potrošači konsultuju sadržaje na ovim sajtovima kako bi sagledali stavove i iskustva drugih potrošača. Predmet onlajn recenziranja varira od tehničkih uređaja (računara, foto-aparata, mobilnih telefona i sl.), preko knjiga, hotela, restorana, automobila, pa do specifičnih usluga poput obrazovanja. Na vebu se dnevno objavi na hiljade korisničkih recenzija, a među njima su i recenzije nastavnog osoblja, koje studenti ostavljaju na sajtovima društvenih mreža i medija i time ih čine javno dostupnim.

Sajtovi društvenih mreža i medija zauzimaju značajno mesto u marketing strategiji obrazovnih institucija na američkom tržištu i u zemljama Zapadne Evrope, dok se u Srbiji njihova upotreba zadržava na otvorenim *Facebook* stranicama ili *Twitter* naložima visokoškolskih institucija. Razmenjeni sadržaj na sajtovima društvenih mreža i medija, onlajn recenzije, kao i podaci o posetama i korisnicima u Srbiji se ne eksploatišu u značajnijoj meri. Istraživanja iz oblasti sentiment analize, do sada pretežno rađena za englesko govorno područje, potvrđuju da sadržaji sa društvenih medija predstavljaju odličan resurs za analizu paterna i trendova, za bolje razumevanje potrošačkih navika i preferencija i proaktivno inoviranje proizvoda ili usluga, što je detaljnije prikazano u [Poglavlju 2](#). Iskustva iz drugih domena navode na zaključak da postoje indicije da bi visoko školstvo takođe imalo koristi od primene sentiment analize nad sadržajima koje studenti generišu, čime bi se omogućilo sagledavanje mišljenja, stavova i osećanja studenata u vezi sa institucijom, nastavnim sadržajima i predmetima, ophođenjem nastavnog i nenastavnog osoblja i drugim aspektima školovanja, što ujedno predstavlja i osnovni predmet istraživanja doktorske disertacije. Kako je reč o interdisciplinarnom problemu koji objedinjuje područje društvenih medija i područje analize podataka (*data mining*, sentiment analiza) u prvom delu uvodnog poglavlja dat je pregled i opis povezanih oblasti istraživanja. Detaljan opis celokupnog toka istraživanja dat je u [Poglavlju 1.3](#), dok [Poglavlje 1.4](#) opisuje specifičnosti opšte terminologije koja se u nastavku rada koristi.

Doktorska disertacija je strukturirana nadalje na sledeći način. [Poglavlje 2](#) daje pregled dosadašnjih praktičnih i teoretskih istraživanja iz oblasti sentiment analize, čime se stiče uvid u široki spektar mogućnosti ovog analitičkog pristupa. [Poglavlje 3](#) opisuje metodologiju sentiment analize, poteškoće u njenoj implementaciji, kao i pregled metoda, tehnika i pristupa, čime su sagledani preduslovi koji moraju biti ispunjeni da bi se pojedini pristupi sentiment analizi realizovali, kao i mogućnosti za njihovu primenu u konkretnom istraživanju. [Poglavlje 4](#).

predstavlja centralni deo doktorske disertacije. U njemu su detaljno opisani metodološki aspekti i tok istraživanja. Osnovni izvor podataka za realizaciju istraživanja, opisan u [Poglavljju 4.1.1](#), predstavljaju onlajn recenzije studenata prikupljene sa sajta *Oceni profesora*. Polazni cilj je predstavljalo identifikovanje načina na koji studenti iskazuju mišljenja o profesorima i njihovo prevođenje u formalni model. Predložen je model prema kojem se u kasnijem koraku vrši anotacija sadržaja, na nivou rečenice i na nivou reči. Implementacija modela je realizovana kroz adekvatnu anotacionu shemu i manuelno označavanje (anotaciju) prikupljenog skupa studentskih recenzija, čime je kreiran relevantan korpus na srpskom jeziku. Navedeni koraci opisani su u [Poglavljima 4.1.3.1–4.1.3.2](#). Na osnovu rezultujućeg korpusa, takozvanog zlatnog standarda, kvantifikovao se i okarakterisao problem kroz statističku eksploraciju korpusa, datu u [Poglavljju 4.1.3.3](#). Korpus je poslužio i za evaluaciju primenjenih algoritama u sentiment analizi. U [Poglavljju 4.1.5](#), opisane su primenjene metode, tehnike i pristupi sentiment analizi nad kreiranim korpusom. Fokus je na identifikovanju ekspresija sentimenta i utvrđivanju njihovog polariteta. Realizovan je niz eksperimenata primenom različitih algoritama mašinskog učenja i sentiment rečnika. Evaluacija razvijenih modela prikazana je u [Poglavljju 4.1.6](#). [Poglavlje 4.2](#) ilustruje mogućnosti primene rezultata sentiment analize u poslovanju visokoškolskih institucija, nakon čega slede zaključna razmatranja i osvrt na istraživačka pitanja koja će biti u fokusu budućih istraživanja.

## **1.1. Društvene mreže i društveni mediji – pojam, mesto i uloga u savremenom društvu**

Društveni mediji predstavljaju platformu za razmenu informacija i participaciju korisnika kako bi se kreirao i/ili distribuirao sadržaj (Steenkamp, et al., 2014). Njihova pojava je rezultat razvoja veb 2.0 tehnologija, koje su omogućile korisnicima da postanu aktivni kreatori sadržaja na vebu i time doprinesu velikoj ekspanziji društvenih medija kao instrumenta onlajn komuniciranja.

Društveni mediji obuhvataju širok spektar sajtova sa jednom zajedničkom osobinom: omogućavaju korisnicima interakciju sa samim sajtom i sa drugim korisnicima. Interakcija može biti jednosmerna, poput komentaranja, ocenjivanja ili glasanja o nekom sadržaju ili kompleksna, poput preporuke proizvoda na osnovu ocena drugih korisnika sajta (Grljević, et al.). Svaki sajt koji poziva korisnike na interakciju može se definisati kao društveni medij. U nastavku su opisani dominantni tipovi sajtova društvenih medija.

**Društveno označavanje** (engl. *social bookmark*). Interakcija među korisnicima se ostvaruje označavanjem veb sajtova i pretragom sajtova koje su drugi korisnici ranije označili. Primeri nekih od vodećih sajtova ove vrste su *Delicious*<sup>1</sup> i *Pinterest*<sup>2</sup>. *Delicious* omogućava jednostavno

---

<sup>1</sup> <https://delicious.com/>

<sup>2</sup> <https://www.pinterest.com/>



označavanje sajtova, čime korisnik formira kolekciju omiljenih veb lokacija na osnovu tagova i ključnih reči, dok kolekciju jednog korisnika drugi posetioci mogu pratiti. *Pinterest* predstavlja sajt za vizuelno označavanje, gde korisnici kreiraju kolekcije omiljenih fotografija sa interneta.

**Društvene vesti** (engl. *social news*). Interakcija se ostvaruje glasanjem za članke i komentarisanjem članaka od strane registrovanih korisnika. Najpoznatiji sajtovi društvenih vesti su *Digg*<sup>3</sup> i *Reddit*<sup>4</sup>. *Digg* nudi širok spektar tema poput tehnologije, zabave i uopštenih tema. *Reddit* je vrlo jednostavnog interfejsa, ali nudi mogućnost praćenja vesti iz bogatog seta tema podeljenih u potkategorije (engl. *subreddits*) u okviru kojih se teme pozicioniraju na vrh stranice na osnovu glasova posetilaca. Takođe, daje mogućnost svakome da kreira zajednicu, ili *subreddit*, koja je nezavisna i kojom upravlja tim volontera.

**Društvena razmena slika i video sadržaja** (engl. *social photo and video sharing*). Interakcija posetilaca sajta je bazirana na razmeni slika i video-zapisa, kao i komentarisanju istih. Predstavnicima ove grupe sajtova su *Instagram*<sup>5</sup> i *YouTube*<sup>6</sup>. *Instagram* pruža mogućnost svojim korisnicima da podele sa zajednicom prijatelja svoj život u slici i filmu, dajući mogućnost njihove obrade kroz specijalizovan softver. *YouTube* je najpoznatiji sajt za deljenje video sadržaja.

**Wiki**. Interakcija korisnika wiki sajtova je bazirana na kreiranju novih članaka i editovanju postojećih. Najpoznatiji wiki sajt je *Wikipedia*<sup>7</sup>. Ona predstavlja višejezičnu, veb baziranu i besplatnu enciklopediju čije sadržaje kreiraju njeni članovi. Broj članaka rapidno raste, te je 2015. godine njihov broj preko 4.900.000.

**Društveno umrežavanje** (engl. *social networking*) je bazirano na određenim strukturama koje omogućavaju korisnicima da iskažu svoju individualnost i povežu se sa ljudima sličnih interesovanja. Ovakve strukture podrazumevaju postojanje korisničkog profila, dodavanje prijatelja, objave u formi bloga, softverskih dodataka (engl. *widgets*) i obično specifičnosti svojstvene konkretnom sajtu društvenog umrežavanja. Većina sajtova društvenog umrežavanja daje mogućnost formiranja i priključenja grupama koje omogućavaju korisnicima da pronađu osobe sličnih interesovanja ili se upuste u diskusije na određenu temu. Najpoznatiji sajtovi ove grupe su *Facebook*<sup>8</sup> i *LinkedIn*<sup>9</sup>. *Facebook* dozvoljava registrovanim korisnicima da kreiraju profile, postave slike i video zapise i šalju poruke. Dostupna je verzija na 37 različitih jezika, a obuhvata i javne platforme poput tržnice gde se razmenjuju oglasi, kreiranje grupa ljudi sa zajedničkim interesovanjem i promotivne stranice. *LinkedIn* je sajt namenjen poslovnom

---

<sup>3</sup> <http://digg.com/>

<sup>4</sup> <https://www.reddit.com/>

<sup>5</sup> <https://instagram.com/>

<sup>6</sup> <https://www.youtube.com/>

<sup>7</sup> <https://www.wikipedia.org/>

<sup>8</sup> <https://www.facebook.com/facebook>

<sup>9</sup> <https://www.linkedin.com/>

umrežavanju, razmeni poslovnih iskustava, oglašavanju slobodnih pozicija u kompaniji i objavi vesti iz poslovanja.

Društveni mediji predstavljaju okruženje u kojem se društveno umrežavanje odvija i koje je izmenilo način na koji potrošači prikupljaju informacije i donose odluke o kupovini. Međutim, usled široke rasprostranjenosti sajtova društvenog umrežavanja i činjenice da i drugi sajtovi društvenih medija ispoljavaju slične karakteristike, poput kreiranja ličnih profila, povezivanja sa drugim korisnicima i komentarisanja njihovih sadržaja, društveni mediji i sajtovi društvenog umrežavanja (u nastavku društvene mreže) se često koriste kao *sinonimi*, kao što je istaknuto u brojnim radovima ((Hester, et al., 2013), (Sumners, 2010), (Paquette, 2013), (Sianipar, et al., 2012), (Ellison, et al., 2013)). Slika 1.1. ilustruje problematiku. Osnovne funkcionalnosti sajtova društvenih medija se ogledaju u objavljivanju i deljenju sadržaja, diskutovanju i komentarisanju sadržaja drugih pojedinaca i umrežavanju. Kako slika prikazuje, određene sajtove društvenih medija karakterišu samo pojedine navedene funkcionalnosti, dok sajtovi društvenog umrežavanja (*Facebook*, *Twitter* ili *Google+*) poseduju sve navedene karakteristike.



Slika 1.1. Spektar sajtova društvenih medija<sup>10</sup>

**Blogovi** predstavljaju diskusione ili informativne sajtove sačinjene iz pojedinačnih poruka u kojima pojedinci ili grupa autora iskazuju mišljenja, stavove i iskustva na određenu temu. Pojava mikroblogova je objedinila blogove pojedinačnih autora i blogove grupe autora u društveni tok vesti. *Twitter*<sup>11</sup> je najpoznatiji mikroblog sajt koji je baziran na principima društvenog umrežavanja. Omogućava registrovanim korisnicima da se povežu sa drugim korisnicima, da postavljaju objave, tzv. tvitove, dužine do 140 karaktera i da odgovaraju na tuđe tvitove.

<sup>10</sup> Preuzeto sa: <http://www.fredcavazza.net/2014/05/22/social-media-landscape-2014/>

<sup>11</sup> <https://twitter.com/>

**Diskusioni forumi** omogućavaju registrovanim korisnicima da rešavaju probleme i nedoumice, razmenjuju informacije, mišljenja i iskustva o određenoj temi u formi iskaza ili pitanja. Forumi su obično fokusirani na specifičnu temu, proizvod ili događaj. Onlajn prodavnice, poput *Amazon*<sup>12</sup> i *eBay*<sup>13</sup>, imaju forume namenjene diskusiji o kupljenim proizvodima i obavljenim transakcijama.

Razvoj sajtova društvenih mreža i društvenih medija značajno je izmenio prirodu ljudskih aktivnosti, interakcija i načina diseminacije informacija. Stvarni društveni odnosi su migrirali u virtuelni svet, što je rezultovalo stvaranjem onlajn zajednica koje povezuju pojedince iz celog sveta i omogućavaju razmenu znanja, zabavu i komunikaciju između različitih kultura (Borges Tiago, et al., 2014). Bihevioralna psihologija sugerise da ljudi imaju urođenu sklonost da saraduju i dele informacije i resurse, očekujući određenu korist (Osatuyi, 2013). U kontekstu društvenih medija, odluka da se određena vrsta informacije podeli može biti zasnovana na koristi koja će se izvesti iz samog procesa deljenja (npr. podizanje reputacije u svojoj onlajn zajednici). Trenutno, *Facebook* ima 1.49 milijardi, *Twitter* 316 miliona, a *Google+*<sup>14</sup> 300 miliona mesečno aktivnih korisnika<sup>15</sup>. Svi ovi korisnici onlajn društvenih mreža i medija su i potrošači različitih proizvoda i usluga koji dnevno razmenjuju informacije, iskustva i mišljenja i čine ih javno dostupnim na veb stranicama. Ovako generisan korisnički sadržaj, koji se pokazao efikasnijim od tradicionalne marketinške komunikacije u pogledu uticaja na stavove i ponašanje korisnika (Thackeray, et al., 2008), predstavlja jednu od osnovnih karakteristika društvenih medija. Pritiskom dugmeta „svida mi se“, odnosno pozitivnim komentarima o proizvodima i uslugama, korisnik utiče na svoje prijatelje da kupuju ili koriste ono što i on. Na ovaj način društvene mreže i mediji, pre svega forumi, blogovi, *Facebook*, *Twitter* i *YouTube* postaju neizostavan izvor informacija koje ljudi koriste pri donošenju konačne odluke o kupovini proizvoda ili usluge, dok u isto vreme predstavljaju efikasnu poslovnu tehniku za kompanije. Značaj koji veb 2.0 aplikacije imaju u savremenom životu i poslovanju naveo je organizacije da razvijaju i održavaju svoje prisustvo na društvenim mrežama i sajtovima društvenih medija kako bi pospešile interesovanje za organizaciju i razvile odnose sa onlajn javnošću (Parveen, et al., 2015). Upotreba društvenih mreža i medija od strane *Fortune 500* kompanija je značajno porasla u 2012. godini. Istraživanje je pokazalo da 73% ovih kompanija ima svoj zvanični nalog na *Twitter*-u, dok 66% ima nalog na *Facebook* stranici. Dodatno, 2011. godine 28% kompanija je posedovalo sopstveni korporativni blog, što je predstavljalo najveće povećanje u upotrebi blogova od 2008. godine (Barnes, et al., 2012). Dugoročni uzlazni trend u upotrebi društvenih medija ukazuje da njihov značaj, kao strategijskog organizacionog alata, raste.

Ponašanje potrošača tokom potrage za informacijama o proizvodima i uslugama je značajno izmenjeno. Oflajn izvori su zamenjeni elektronskim marketingom od usta do usta (engl. *Word of Mouth* - eWOM), poput društvenih mreža ili sajtovima za recenziranje (Gruen, et al., 2006). Pri

---

<sup>12</sup> <http://www.amazon.com/>

<sup>13</sup> <http://www.ebay.com/>

<sup>14</sup> <https://plus.google.com/>

<sup>15</sup> Podaci su preuzeti sa zvaničnih sajtova Facebook, Twitter i Google+. Svaki profil, grupa, ili stranica se tretiraju kao korisnici.

tome, potrošači su pokazali da više cene mišljenje drugih pojedinaca i veruju im više nego promotivnim kampanjama kompanije ((Berthon, et al., 2012), (Pitt, et al., 2002)) i da ukazuju jednako poverenje u onlajn kritike i u lične preporuke prijatelja ((Park, et al., 2007), (Gligorijevic, et al., 2012)). Ekspanzija upotrebe društvenih mreža i medija je dovela do promene u lokusu kontrole. Onlajn kritike se nalaze na trećem mestu po uticaju koji imaju na odluke o kupovini, nakon kupona i popusta ((Yang, et al., 2015/2), (Marketing\_Charts, 2014)). Pored uticaja koji imaju na stavove i akcije potrošača, one povećavaju stepen poverenja potrošača u kompaniju, direktno utiču na potrošačev odabir preduzeća među konkurentima i utiču na uspešnije privlačenje novih potrošača, te je od izuzetne važnosti kako će kompanija reagovati na pritužbe i pohvale svojih potrošača i kako će upravljati svojom onlajn reputacijom jer bi negativne kritike mogle imati snažne posledice na poslovanje – ugroziti privlačenje novih potrošača, lojalnost postojećih potrošača, uticati na visinu prihoda i drugo.

Usled promena nastalih u ponašanju potrošača, kompanije su prihvatile društvene mreže i medije i promenile način komunikacije. Iz ugla kompanija, najzastupljeniji motivi upotrebe društvenih medija su jačanje poverenja, prepoznatljivost proizvodne marke i uticaj na opredeljenje potrošača (Dijkmans, et al., 2015). Aktivnosti usmerene ka ostvarivanju ovih ciljeva se nazivaju upravljanjem onlajn reputacijom koja podrazumeva onlajn interakcije sa potrošačima, kreiranje sadržaja koji se može deliti, praćenje mišljenja potrošača i njihovih dijaloga, rešavanje problema prepoznatih u negativnim sadržajima podeljenim na internetu, kao i poštovanje iskazanih ideja i sugestija potrošača.

U radu (Dijkmans, et al., 2015) autori su sumirali razloge koji čine korporativnu reputaciju relevantnom. U procesu odabira dobavljača ključni parametar koji potencijalni potrošači razmatraju jeste reputacija kompanije. Potrošači češće odabiraju kompanije sa pozitivnom reputacijom i spremni su da plate više za njihove proizvode. Pozitivna reputacija može stvoriti tržišne barijere za konkurenciju, doprineti negovanju lojalnosti potrošača i njihovom zadržavanju i vezivanju za kompaniju, kao i omogućiti kompanijama da privuku veći broj novih potrošača, što posledično rezultuje povećanjem tržišne vrednosti. Takođe, pozitivna korporativna reputacija štiti kompaniju u kriznim uslovima i povećava spremnost interesnih grupa da investiraju u kompaniju.

Angažovanost kompanije na sajtovima društvenih mreža i medija jeste korisna za reputaciju, međutim, postoje određeni nedostaci. Društvene mreže i mediji ne predstavljaju jednosmerni kanal komunikacije, već nekontrolisanu arenu za učesnike, što može predstavljati opasnost i dovesti do narušavanja ugleda kompanije. Korisnici slobodno iskazuju svoja mišljenja i ideje koje mogu biti u sukobu sa poslovnim interesima kompanije. Usled brzine širenja informacija kroz društvene medije, čak i jedan nezadovoljan potrošač može urušiti reputaciju kompanije (Tripp, et al., 2011). Uslužne kompanije su podložnije ovom uticaju od drugih zbog prirode njihovih proizvoda. Usluge su nematerijalne, nestandardizovane i moraju se konzumirati kako bi mogle da se evaluiraju, što povećava mogućnost pojave raskoraka između očekivanja kupaca,

njihove percepcije i same usluge, te posledično može dovesti do učestalijeg iznošenja pritužbi na sajtovima društvenih medija (Dijkmans, et al., 2015).

Stepen u kom su potrošači angažovani u aktivnostima na društvenim medijima kompanije je odraz njihove percepcije o reputaciji kompanije. Angažovanost se definiše kao kombinacija kognitivnog aspekta (npr. potrošač je zainteresovan za aktivnosti kompanije), bihevioralnog aspekta (participacija u aktivnostima) i/ili emocionalnog aspekta (pozitivan stav prema aktivnostima kompanije) (Dijkmans, et al., 2015). Postizanje visokog stepena angažovanosti potrošača pospešuje reputaciju kompanije i lojalnost prema brendu i utiče na odluke o kupovini. Autori u radu (Hudson, et al., 2015) ističu relevantnost emocija u procesu gradnje veze sa potrošačima i brendom, kao i njihov uticaj na spremnost potrošača da preporuči proizvod.

Emocije predstavljaju pozitivna (draga, relaksirajuća) ili negativna (nervozna, ljutita) osećanja osobe. Autori u radu (Hudson, et al., 2015) sumiraju istraživanja o načinu procesiranja emocija koja potvrđuju da se pre bilo kakve obrade informacija formiraju emocije, te da je i ponašanje potrošača pod njihovim snažnim uticajem. U radu (Pawle, et al., 2006) autori su otkrili da su faktori sa najsnažnijim sveukupnim uticajem na kupovne namere pojedinca emocionalni faktori koji pokreću ljubav, pogotovo intimnost praćena misterijom, koja počiva na čulnim zadovoljstvima. Autori sugerišu da je ključni emocionalni triger za jačanje veze između potrošača i brenda stvaranje snažne intimne povezanosti sa brendom (brend je relevantan za potrošača, potrošač oseća povezanost sa brendom i strastven je u vezi sa brendom). Usled činjenice da danas potrošači reaguju pozitivnije na sadržaj podeljen na društvenim medijima, nego na klasične reklamne kampanje, njihov emocionalni odgovor na ovakve interakcije jača vezu prema brendu. Takođe, kada se potrošačima daje mogućnost da ostvare interakcije sa kompanijom na društvenim medijima kroz deljenje znanja o brendu, potrošačko iskustvo ili sugestije, jača se veza potrošača sa brendom i kompanijom. Očekivanja koja potrošači imaju vode ka emocionalnom odgovoru rezultujući pozitivnim ili negativnim osećanjem koje oblikuje njihovo zadovoljstvo, poverenje i lojalnost. Otuda, adekvatnom analizom sadržaja dostupnih na sajtovima društvenih medija, a prevashodno analizom emocija skrivenih u korisnički generisanom sadržaju, kompanije mogu dobiti kritične informacije za unapređenje svog poslovanja. Uvođenjem potrošačke perspektive u usluge i poslovne transakcije, organizacije mogu isporučiti potrošačima proizvod ili uslugu koja je najviše prilagođena njihovim potrebama.

## **1.2. Analiza podataka o mišljenjima i sentiment**

Značajan porast javno dostupnih sadržaja na internetu u kojima se na eksplicitan ili implicitan način iskazuje mišljenje nameće sve veću potrebu za njihovom ekstrakcijom i analizom kako bi se stavili u funkciju unapređenja poslovanja, za razumevanje konzumentskih izbora, namera i sentimenta u stvarnom vremenu. Velikom količinom podataka se teško upravlja i njihova pretraga zahteva dosta vremena, te je značajna uloga u potrazi za korisnim i kvalitetnim

informacijama kroz adekvatno filtriranje, eksploataciju i analizu pripala metodama i tehnikama *data mining*-a (Liu, 2012). Njihova primena na sadržaje sa sajtova društvenih medija je objedinjena pod nazivom analiza društvenih medija (engl. *social media analytics* – SMA) koja omogućava kompanijama da evaluiraju prikupljene podatke sa sajtova društvenih medija, kako bi se upoznali sa bazom svojih konzumenata na višem nivou – koji će uključiti i analizu njihovih emocija, sentimenta i mišljenja, što će posledično omogućiti optimizaciju marketing poruke, pozicioniranje brenda i razvoj proizvoda ili usluge. Osnovni ciljevi i benefiti primene SMA se ogledaju u povećanju profita, smanjenju troškova vezanih za podršku potrošačima, prikupljanju povratnih informacija o uslugama i proizvodima, kao i poboljšanju mišljenja javnosti o kompaniji ili konkretnom proizvodu ili usluzi. Jedan od aspekata SMA predstavlja analiza osećanja, takozvana sentiment analiza (engl. *sentiment analysis*) ili, kako se često u literaturi naziva, *opinion mining*. Iako se termini sentiment analiza i *opinion mining* koriste kao sinonimi, pojedini autori smatraju da je reč o donekle različitim pojmovima. Sentiment analiza predstavlja istraživačku oblast koja se bavi analizom mišljenja javnosti, sentimenta, stavova i emocija usmerenih ka određenom entitetu. Entitet može biti pojedinac, događaj ili određena tema. Na sajtovima društvenih medija teme se obično obrađuju kroz recenzije korisnika. Sentiment analiza se opisuje u (Nasukawa, et al., 2003) kao snimanje naklonosti, u (Yi, et al., 2003) kao ekstrahovanje sentimenta o određenoj temi. Sentiment analiza identifikuje iskazane sentimente u tekstovima, a potom ih analizira, dok *opinion mining* ekstrahuje i analizira mišljenja pojedinaca o određenom entitetu. Iz toga proizilazi da je cilj sentiment analize da pronađe mišljenja, identifikuje sentiment koji je iskazan i izvrši njegovu klasifikaciju prema polaritetu (Medhat, et al., 2014). Usled razlika u definicijama i konceptima, predmetna terminologija je preciznije definisana u [Poglavlju 1.4](#).

Istorija razvoja *opinion mining*-a i sentiment analize je relativno kratka, jer je pre pojave *World Wide Web*-a i intenziviranja upotrebe interneta bilo malo javno dostupnog teksta obojenog mišljenjima. Klasifikacija sentimenta u kritikama proizvoda na one koji izražavaju pozitivan i negativan sentiment se pojavljuje 2003. godine u radu (Dave, et al., 2003) što ujedno predstavlja jedan od većih iskoraka u ovoj oblasti. Veliki deo istraživanja sprovedenih nakon ovog rada izvodi iste zaključke, pre svega da *opinion mining* alati treba da ekstrahuju i analiziraju mišljenja o različitim aspektima proizvoda ili usluga.

Iako lingvistika i obrada prirodnog jezika (u tekstu NLP – engl. *Natural Language Processing*) imaju dugu istoriju, pre 2000. godine malo istraživanja je rađeno o iskazanim sentimentima i ljudskom mišljenju kroz tekstove u slobodnoj formi. Nakon toga, sentiment analiza postaje aktivna istraživačka oblast. Autor u (Liu, 2012) ističe sledeće razloge. Primarno, sentiment analiza ima široku primenu u gotovo svim domenima. Industrije koje okružuju sentiment analizu takođe su napredovale usled razvoja komercijalnih aplikacija i njihove primene, što pruža snažnu motivaciju za dalja istraživanja. Zatim, sentiment analiza nudi mnoštvo novih izazovnih istraživačkih problema. Kao presudni razlog intenziviranja istraživanja u oblasti sentiment analize autor navodi činjenicu da prvi put u ljudskoj istoriji postoji velika količina podataka

obojenih mišljenjima, koja su javno dostupna na društvenim medijima i webu. Bez ovih podataka, brojna istraživanja ne bi bila moguća. Otuda, nastanak i brz razvoj sentiment analize poklapa se sa nastankom i razvojem društvenih mreža i medija, a sentiment analiza trenutno zauzima centralno mesto u istraživanjima iz oblasti društvenih medija. Slika 1.2. ilustruje kretanje interesovanja ljudi širom sveta za pojmove “*sentiment analysis*” (plava linija) i “*social media analytics*” (crvena linija). Može se zapaziti da je pretraga ovih pojmova putem *Google* internet pretraživača započeta početkom 2008. godine, odnosno u prvoj polovini 2009. godine. Iako postoje oscilacije u interesovanju javnosti generalni rastući trend se zadržao do 2015. godine. Istraživanja iz oblasti sentiment analize imaju snažan uticaj na NLP, ali takođe i na menadžment, političke, ekonomske i društvene nauke, jer su sve one pod uticajem mišljenja javnosti.



Slika 1.2. Nivo interesovanja tokom vremena, Google Trends<sup>16</sup>

U osnovi sentiment analize se nalaze metode i tehnike rudarenja podataka, odnosno *data mining*-a. *Data mining* predstavlja jedan od najboljih načina za obradu i analizu velike količine podataka dostupnih na internetu, jer omogućuje objektivnu karakterizaciju društvenog ponašanja pojedinaca, grupe ili organizacije (Yi, et al., 2003). Koncept *data mining*-a se odnosi na analizu istorijskih podataka sa ciljem otkrivanja šablona u njihovim kretanjima, koji će omogućiti izgradnju prediktivnih modela ili sticanje saznanja o korisnicima. *Data mining* tehnika, koja je u osnovi najvećeg broja *opinion mining* aplikacija, jeste klasifikacija, uključujući regresiju i rangiranje. Većina istraživanja iz ove oblasti se bavi problemom klasifikacije teksta prema polaritetu ili orijentaciji iskazanog mišljenja na pozitivne, negativne i eventualno neutralne kritike.

U svakom *data mining* poduhvatu na pripremu podataka odlaze najveći naponi i ova faza troši najviše vremena analitičara. Mišljenja i osećanja dostupna u podacima sa interneta nose svoje

<sup>16</sup> *Google Trends* su dostupni na sajtu:

<https://www.google.rs/trends/explore#q=sentiment%20analysis%2C%20social%20media%20analytics&cmp=q&tz=Etc%2FGMT-1>

specifičnosti koje dodatno otežavaju *data mining* proces. Pisana su na prirodnom jeziku te je reč o nestrukturiranim podacima koji često uključuju polaritet, različitu orijentaciju sentimenta i gradacije (McGlohon, et al., 2010). Dodatno, svaki jezik nosi svoje specifičnosti, te je u razvoju aplikacija sentiment analize važno posvetiti se i specifičnostima jezika na kom su tekstovi pisani i načinu na koji se one odražavaju na pripremu podataka.

Objektivna rečenica iskazuje činjenicu, dok subjektivna rečenica iskazuje osećanje, mišljenje ili uverenje. Prisutnost subjektivnosti u rečenicama je snažno povezana sa domenom. Kako je naglašeno u (Read, 2005), zavisnost od domena je delimično posledica promene u vokabularu, te identični izrazi mogu nositi potpuno drugačiji sentiment u različitim domenima (npr. nalog „Pročitaj knjigu“ se može tretirati kao pozitivan sentiment kada je reč o kritikama knjige, a negativan kada je reč o kritikama filma). Dodatno, klasifikator obučen skupom kritika o jednom tipu proizvoda često ne pokazuje istu performantnost kada se primeni nad skupom kritika o drugom proizvodu (Feldman, et al., 2011). Zbog evidentne domenske zavisnosti, da bi se sentiment analiza mogla realizovati, potrebno je kreirati resurse za dati domen, prevashodno anotirane korpuse. Kao rezultat postupka anotacije dobijaju se označeni podaci koji omogućavaju primenu različitih algoritama mašinskog učenja, pre svega nadgledanih klasifikacionih algoritama, kao i njihovu evaluaciju.

Izazovi sentiment analize su detaljnije opisani u [Poglavlju 3.2](#), dok su aplikacije i primene sentiment analize prikazane u [Poglavlju 2](#). U nastavku su prikazani ciljevi i postupak istraživanja koje je predmet doktorske teze i jasno je razgraničena predmetna terminologija upotrebljena u radu.

### 1.3. Opis istraživanja i pristupa

Na teritoriji Republike Srbije upotreba društvenih mreža i medija u potrošačkom društvu predstavlja neistraženu oblast, pa ne možemo argumentovano govoriti o stanju u našoj zemlji: da li pratimo svetske tendencije, da li su navike konzumenata iste ili drugačije, da li su preduzeća u Srbiji svesna velikih uticaja elektronskog marketinga „od usta do usta“ i u kojoj meri i za koje namene ga koriste. Zvanični podaci Republičkog zavoda za statistiku<sup>17</sup> ukazuju na značajan porast broja domaćinstava sa računarom i internet priključkom, te da se internet prvenstveno koristi za posetu društvenim mrežama i medijima i pretragu informacija o obrazovanju i kursevima. Ovakav patern upotrebe interneta ukazuje na mogućnost otvaranja obrazovnog sektora ka korisnicima putem kanala društvenih medija.

Privlačenje studenata predstavlja jedno od ključnih pitanja rada i opstanka visokoškolskih institucija u Republici Srbiji zbog veoma izraženog demografskog pada, koji je doveo do

---

<sup>17</sup> Republički zavod za statistiku Republike Srbije, Upotreba informaciono-komunikacionih tehnologija u Republici Srbiji 2013, dostupno na: <http://webrzs.stat.gov.rs/WebSite/repository/documents/00/01/14/03/PressICT2013.pdf>



fenomena poznatog kao „bela kuga“. Naime, usled značajnog pada nataliteta u Srbiji, generacija srednjoškolaca koja će 2020. upisivati fakultet brojaće gotovo isto toliko đaka koliko će biti raspoloživih mesta za upis na visokoškolske institucije<sup>18</sup>, što znači da predstoji „borba“ za svakog studenta i samo one institucije koje uspeju da privuku zadovoljavajući broj studenata moći će uspešno poslovati. Međutim, naučna zajednica u Srbiji se nije bavila detaljnijom analizom upotrebe društvenih medija u visokom obrazovanju, niti sentiment analizom sadržaja o visokoškolskim institucijama, te pred fakultetima tek stoji mogućnost iskorišćenja benefita ovih analiza kako bi uspešnije targetirali i privukli nove studente, pratili stepen zadovoljstva postojećih studenata, potencijalno identifikovali uska grla, pratili tržišne trendove i iskoristili ova saznanja za unapređenje usluga. Takođe, iz perspektive univerziteta, bilo bi korisno istražiti uticaj javnog raspoloženja na izbor pojedinih fakulteta ili studijskih programa.

Jedna od osnovnih prepreka u razvoju sentiment analize u Srbiji jeste nedostatak anotiranih korpusa za različite domene. Kreiranje anotiranog korpusa za domen visokog obrazovanja predstavlja centralni deo istraživanja doktorske disertacije, čime će se stvoriti potrebna baza za istraživanje primene sentiment analize nad sadržajima pisanim na srpskom jeziku. Nadalje, nedostaje detaljno definisan i sistematizovan okvir koji će voditi istraživanje u oblasti analize sadržaja, od odabira skupa, pripreme podataka, do primene adekvatnih algoritama u skladu sa ciljem analize. Nema mnogo studija i istraživanja koja se bave proučavanjem primene *data mining* metoda i tehnika specifičnih za pojedinačne faze sentiment analize, a kroz istraživanje koje je predmet doktorske disertacije analizira se primenljivost različitih algoritama *data mining*-a kako bi se utvrdio najadekvatniji za sentiment analizu srpskog jezika nad pripremljenim korpusom.

Definisani predmet istraživanja ove disertacije otvara brojne *istraživačke probleme i pitanja* (IP) iz čijeg skupa su izdvojena sledeća:

**IP1.** Sentiment analiza predstavlja mladu istraživačku oblast. Iako su dostignuća na engleskom govornom području raznolika, mnogi problemi još uvek nisu rešeni zbog kompleksnosti oblasti i brojnih nametnutih izazova. Stoga ne iznenađuje što su generalno u Srbiji istraživanja iz oblasti analize sadržaja sa društvenih medija neznatna, a u potpunosti izostaju u oblasti visokog obrazovanja. Iz ovoga proizilazi prvi istraživački problem: na koje načine, za koje svrhe i potrebe mogu da se rade analize sadržaja društvenih medija, i kako se dobijeni rezultati mogu iskoristiti za unapređenje poslovanja visokoškolskih institucija?

**IP2.** S obzirom na zastupljenost engleskog jezika u društvenim medijima u vidu brojnih statusa i komentara, a pogotovo značaja koji ovaj jezik ima u naučnoj zajednici, razumljivo je da je i najveći broj analitičkih pristupa i primenjenih tehnika i metoda analize sadržaja razvijen upravo za englesko govorno područje. Stoga se

---

<sup>18</sup> Strategija razvoja obrazovanja u Srbiji do 2020. godine („Sl. glasnik RS“, br. 107/2012), dostupna na: [http://www.uns.ac.rs/sr/novosti\\_dogadjaji/docs/strategijaRazvojaObrazovanja2020.pdf](http://www.uns.ac.rs/sr/novosti_dogadjaji/docs/strategijaRazvojaObrazovanja2020.pdf)

otvara istraživački problem primenljivosti razvijenih pristupa pripreme podataka i analitičkih metoda analize sadržaja sa društvenih medija nad sadržajima na srpskom jeziku i problem potrebnih prilagođavanja.

Polazna *hipoteza* doktorske teze je da adekvatno ekstrahovanje i priprema sadržaja o visokoškolskim institucijama sa sajtova društvenih medija i njihova konsekvantna analiza (analiza sentimenta i mišljenja) mogu ukazati na skrivene šablone ponašanja korisnika sajtova. Ovako derivirana saznanja doprinose poboljšanju odnosa visokoškolske institucije i njenih studenata kroz privlačenje većeg broja novih studenata i građenje lojalnosti postojećih, negovanju reputacije i prepoznatljivosti brenda, kao i prilagođavanju usluga preferencijama studenata i tržišnim trendovima.

Glavni *cilj disertacije* (u oznaci C) se može formulisati na sledeći način:

**C1.** Sagledati moguće načine unapređenja poslovanja visokoškolskih institucija oslanjajući se na rezultate analiza sadržaja društvenih medija.

Iz ovog osnovnog, proizilaze sledeći *podciljevi*:

**C1.1.** Utvrditi primenljivost metoda, tehnika i pristupa analize sadržaja sa društvenih medija na oblast visokog obrazovanja, kao i moguće efekte koje bi njihova primena dala na usluge i servise koje visokoškolske institucije nude.

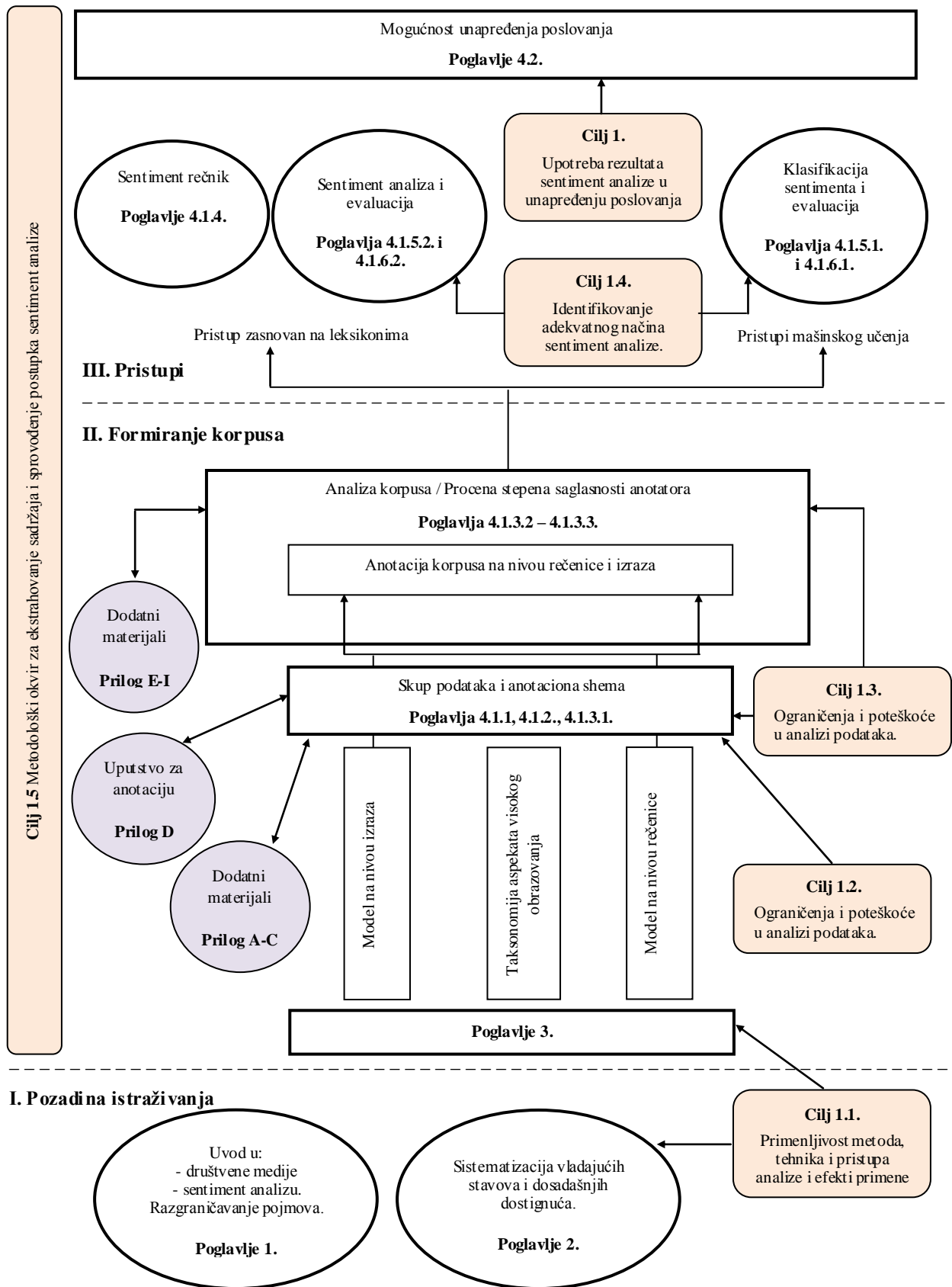
**C1.2.** Identifikovati ograničenja i poteškoće u analizi podataka koje se mogu pojaviti kao posledica specifičnog predmeta analize i usled specifičnosti jezika i ukazati na potencijalna rešenja.

**C1.3.** Formirati anotirani korpus kao osnovu za sprovođenje sentiment analize pominjanja visokoškolskih institucija.

**C1.4.** Identifikovati najadekvatniji način za sentiment analizu srpskog jezika, odnosno sentiment analizu pominjanja visokoškolskih institucija, kroz sistematičnu analizu i poređenje odabranih algoritama.

**C1.5.** Izraditi metodološki okvir za ekstrahovanje sadržaja o pominjanjima visokoškolskih institucija Republike Srbije na društvenim medijima i sprovođenje postupka sentiment analize za potrebe upravljanja reputacijom.

Slika 1.3. ilustruje strukturu doktorske teze i poziciju definisanih ciljeva istraživanja. Jedan od rezultata disertacije će biti, pre svega, jasno razgraničavanje pojmova – [Poglavlje 1.4](#), i sistematizacija vladajućih stavova, kao i dosadašnjih dostignuća iz oblasti analize društvenih medija – [Poglavlje 2](#).



Slika 1.3. Struktura istraživanja

Da bi se analizirao sentiment studenata u pogledu visokoškolskih institucija, potrebno je kreirati osnovni skup podataka kao input za analizu, odnosno korpus. U istraživanju populaciju čine relevantni sadržaji, a u uzorački okvir ulaze sadržaji samo sa određenih sajtova društvenih medija. Uzorački okvir je dovoljno velik jer ga čine najreprezentativniji sajtovi društvenih medija, koji obuhvataju najveći deo sadržaja za definisani predmet istraživanja, i stoga se okvir može izjednačiti sa populacijom. Ostali sajtovi društvenih medija, koji nisu ušli u uzorački okvir, veoma su mali i nisu toliko relevantni za posmatranje. Prema dosadašnjem uvidu u sadržinu društvenih medija, različiti diskusioni forumi i sajt za rangiranje baziran na principima društvenih mreža predstavljaju dominantan izvor podataka sa pominjanjem visokoškolskih institucija u Republici Srbiji. Nakon definisanja uzoračkog okvira, koristiće se slučajni uzorak dobijen preuzimanjem relevantnih sadržaja, značajnih za predmet istraživanja. Na ovaj način će se stvoriti uslovi za davanje odgovora na postavljena istraživačka pitanja. Drugim rečima, izvršiće se analiza informacija i razmenjenih mišljenja o visokoškolskim institucijama, pisanih na srpskom jeziku i prikupljenih sa javno dostupnih izvora na internetu, dakle, bez ograničenja u pogledu geografske lokacije kreatora sadržaja. Disertacija se bavi sistematizacijom koraka prikupljanja i pripreme podataka sa reprezentativnih sajtova, koji uključuju identifikovana ograničenja i poteškoće, kao i njihova potencijalna razrešenja, do ekstrahovanja znanja, kao što je opisano u [Poglavljju 4.1.1](#). Prikupljeni i pripremljeni podaci predstavljaju osnovu za kreiranje anotiranog domensko-specifičnog korpusa, koji trenutno u potpunosti nedostaje za srpsko govorno područje. Postupak kreiranja anotiranog korpusa opisan je u [Poglavljju 4.1.3](#). Nad anotiranim korpusom primeniće se metode i tehnike *data mining*-a čija su teoretska polazišta opisana u [Poglavljju 3.3](#). Ovaj korak daje odgovor na pitanje koje metode i tehnike inteligentne analize podataka su najadekvatnije i najefikasnije za sentiment analizu sadržaja na srpskom jeziku i ukazuje na metodološke aspekte o kojima je neophodno voditi računa u postupku *data mining*-a, s obzirom na domenske i jezičke specifičnosti, kao što je opisano u [Poglavljju 4.1.5](#). Na osnovu rezultata sprovedenog postupka sentiment analize nad ekstrahovanim sadržajima o pominjanjima visokoškolskih institucija u Srbiji na društvenim medijima oblikovaće se preporuke za unapređenje poslovanja, kvaliteta pruženih usluga i odnosa sa studentima prikazanih u [Poglavljju 4.2](#).

## 1.4. Predmetna terminologija

Usled dvosmislenosti pojedinih pojmova i pristupa na koje se istraživanje najviše oslanja, oni su sažeti i formalno definisani u ovom poglavlju. Formalizovanje problema sentiment analize omogućava uvid u bogat skup međusobno povezanih potproblema koji je sačinjavaju. Izučavana literatura bavi se računarskom obradom mišljenja, sentimenta i subjektivnosti. Pored naziva sentiment analiza i *opinion mining*, u literaturi se susreću i drugi nazivi poput ekstrahovanja mišljenja (engl. *opinion extraction*), analiza subjektivnosti (engl. *subjectivity analysis*), analiza recenzija (engl. *review mining*) i drugi. Proliferacija termina odražava razlike u značenju svakog

termina, kako u njihovoj upotrebi u dijalogu, tako i u stručnoj literaturi. Autori u (Pang, et al., 2008) ističu ovu višeznačnost dajući listu definicija usko povezanih sa pojmom osećanja.

- *Mišljenje* (engl. *opinion*) podrazumeva izvedeni zaključak još uvek otvoren za diskusiju („Čini se da je svaki ekspert imao drugačije mišljenje.“).
- *Gledište* (engl. *view*) sugerise subjektivno mišljenje („Veoma je asertivan u navođenju svojih gledišta.“).
- *Uverenje* (engl. *belief*) implicira namerno prihvatanje i saglasnost („Čvrsto verovanje u platformu njene stranke.“).
- *Ubeđenje* (engl. *conviction*) se odnosi na čvrsto i ozbiljno negovanje ubeđenja („Ubeđenje da je život životinje jednako vredan kao i život čoveka [...].“).
- *Osvedočenje* (engl. *persuasion*) sugerise uverenje zasnovano na dokazu ili drugoj potvrdi njegove istinitosti („Ubeđen je da se sve menja.“).
- *Sentiment* sugerise stabilno mišljenje koje reflektuje nečija osećanja („Njena feministička osećanja su dobro poznata.“).

Sentiment analiza se bavi izučavanjem mišljenja koje iskazuje ili ukazuje na pozitivan, odnosno negativan sentiment. Za razliku od činjenica, mišljenja i osećanja su izrazito subjektivna. Iz tog razloga je potrebno analizirati skup mišljenja različitih ljudi umesto pojedinačnog mišljenja koje iskazuje subjektivan pogled pojedinca. Tako sentiment analiza može agregirati sveukupno javno mišljenje koristeći se tehnikama sumiranja mišljenja, i okarakterisati varijacije u naklonostima tokom određenog vremenskog perioda.

Recenzija može biti sačinjena iz jednog ili više iskazanih mišljenja pozitivne ili negativne orijentacije. Mišljenja mogu biti usmerena ka proizvodu/usluzi kao celini ili različitim pojedinačnim aspektima. Dakle,

*Mišljenje je sačinjeno iz dve osnovne komponente: cilja  $g$  i sentimenta  $s$  usmerenog ka cilju, odnosno ( $g, s$ ), gde  $g$  može biti bilo koji entitet ili njegov aspekt o kojem je mišljenje iskazano, a  $s$  je pozitivan, negativan ili neutralan sentiment, odnosno numerička ocena kojom se iskazuje intenzitet sentimenta (npr. skala sa 1-5 zvezdica). Pozitivan, negativan i neutralan sentiment se nazivaju orijentacijom sentimenta ili mišljenja, odnosno polaritetom iskazanog mišljenja (Liu, 2012).*

Na primeru digitalnih foto-aparata, određeni model foto-aparata bi predstavljao entitet, odnosno cilj sentimenta, dok bi njegove karakteristike poput kvaliteta slike, dužine trajanja baterije i sl. bili aspekti entiteta foto-aparat, koji takođe mogu predstavljati cilj iskazanog sentimenta. U kontekstu visokog obrazovanja, jedan profesor bi predstavljao entitet, odnosno cilj sentimenta, dok bi njegove karakteristike – stil predavanja, odnos prema studentima, ljudske karakteristike i vrednosti, predstavljale aspekte entiteta profesor. Cilj iskazanog sentimenta se u literaturi još naziva i temom.

U jednoj recenziji se može naći mišljenje više osoba, takozvanih *nosilaca mišljenja*, npr. kada jedna osoba upoređuje svoje mišljenje sa mišljenjem druge osobe. *Primer A:* „Foto-aparat imam pola godine i obožavam ga. Međutim, moja sestra smatra da je suviše težak za nju.“ *Primer B:* „Kolege su širile negativno mišljenje o profesoru. Međutim, videla sam da je profesor izuzetno prijatan i dobro objašnjava.“ Takođe, obično se uz iskazano mišljenje čuvaju datum i vreme kada je mišljenje iskazano. Datum je u praksi bitan za praćenje varijacija u mišljenju tokom vremena.

Na osnovu iznetog autor je u radu (Liu, 2012) definisao *mišljenje* na sledeći način:

*Mišljenje predstavlja četvorku (g, s, h, t), gde g predstavlja cilj mišljenja ili sentimenta, s je sentiment iskazan ka g, h je nosilac mišljenja, a t označava vreme kada je mišljenje iskazano.*

Ovakva definicija se teško može koristiti u praksi, pogotovo u domenu onlajn recenzija proizvoda i usluga, gde potpuni opis cilja može biti kompleksan i može biti iskazan u više rečenica. Npr.: „Slušao sam predmet kod profesora X u prethodnom semestru. Kvalitet predavanja je na najvišem nivou.“ U drugoj rečenici iskazano je mišljenje o kvalitetu predavanja profesora X, međutim, u rečenici se pominje samo „kvalitet predavanja“. U ovom slučaju cilj recenzije nije samo kvalitet predavanja, jer bez saznanja da se evaluira konkretan profesor mišljenje iskazano u drugoj rečenici nema veliku vrednost. U praksi se cilj iskazanog mišljenja može dekomponovati i opisati na strukturiran način sa višestrukim nivoima, što umnogome olakšava i samu analizu mišljenja i kasniju upotrebu dobijenih rezultata. U prethodnom primeru, profesor X predstavlja entitet, dok kvalitet predavanja predstavlja njegov atribut. U radu (Liu, 2012) su definisali *entitet* na sledeći način:

*Entitet e može biti proizvod, usluga, tema, problem, osoba, organizacija ili događaj. Entitet je opisan parom e: (T, W), gde je T hijerarhija delova, podeoka itd., a W je set atributa entiteta e. Svaki deo takođe ima svoj set atributa.*

U primeru foto-aparata, Canon 600d je entitet koji ima set atributa (kvalitet slike, težina, veličina) i set delova (baterija, objektiv, okular). Baterija takođe ima svoj set atributa (dužina trajanja, težina). U primeru visokog obrazovanja, profesor je entitet koji ima set atributa (kvalitet predavanja, odnos prema studentima). Kvalitet predavanja ima svoj set atributa (dužina trajanja, jasnoća izlaganja, interesantnost). U osnovi, definicija opisuje hijerarhijsku dekompoziciju entiteta na osnovu relacija njenih delova (*part-of* odnosa). Centralni čvor čini naziv entiteta – profesor X/Canon 600d, svi ostali čvorovi hijerarhije su delovi ovog entiteta. Mišljenje se može iskazati o bilo kojem čvoru i bilo kojem atributu. Ovako postavljen entitet sa hijerarhijom sačinjenom od proizvoljnog broja nivoa je zahtevan za reprezentaciju NLP tehnikama i često suvišan (Liu, 2012). Iz tog razloga je uvedena hijerarhija sačinjena iz dva nivoa, kao i termin *aspekt* kojim su objedinjeni atributi i delovi entiteta. Ovako pojednostavljen okvir, gde centralni čvor predstavlja entitet, a drugi nivo su aspekti entiteta, predstavlja radni okvir koji se tipično

koristi u sistemima sentiment analize (Liu, 2012). U literaturi se entiteti često nazivaju objektima, a aspekti karakteristikama (engl. *feature*).

Na osnovu iznetog, u radu (Liu, 2012) redefinisana je definicija mišljenja na sledeći način:

*Mišljenje je petorka  $(e_i, a_{ij}, s_{ijk}, h_k, t_l)$ , gde  $e_i$  predstavlja naziv entiteta,  $a_{ij}$  je aspekt entiteta  $e_i$ ,  $s_{ijk}$  je iskazan sentiment o aspektu  $a_{ij}$  entiteta  $e_i$ ,  $h_k$  je nosilac mišljenja i  $t_l$  je vreme kada je  $h_k$  iskazao mišljenje. Sentiment  $s_{ijk}$  je pozitivan, negativan ili neutralan ili je iskazan različitim stepenima intenziteta, npr. 1-5 zvezdica koje najčešće koriste sajтови za recenziranje. Kada je mišljenje iskazano o entitetu kao celini, koristi se poseban OPŠTI aspekt. Ovde,  $e_i$  i  $a_{ij}$  zajedno predstavljaju cilj mišljenja.*

Na osnovu definicije mišljenja formalizovan je osnovni zadatak sentiment analize na sledeći način (Liu, 2012):

*Ako je dat dokument sa iskazanim mišljenjem  $d$ , otkriti sve petorke mišljenja  $(e_i, a_{ij}, s_{ijk}, h_k, t_l)$  u  $d$ .*

Ključni zadaci sentiment analize su izvedeni iz komponenti definisane petorke. Ukoliko je dat set dokumenata  $D$  u kojima je iskazano mišljenje, zadaci su:

1. Ekstrahovanje svih izraza entiteta (engl. *entity expressions*) i njihovo grupisanje u klastere entiteta, odnosno kategorije entiteta (engl. *entity category*). **Kategorija entiteta** predstavlja jedinstven entitet, dok **izraz entiteta** predstavlja reč ili frazu koja se pojavljuje u tekstu i ukazuje na kategoriju entiteta. Svaki klaster izraza entiteta ukazuje na jedinstven entitet  $e_i$ .
2. Ekstrahovanje svih aspekata definisanih entiteta i njihovo kategorizovanje u klastere, odnosno kategorije. **Kategorija aspekta** (engl. *aspect category*) određenog entiteta predstavlja jedinstveni aspekt entiteta, dok iskaz aspekta (engl. *aspect expression*) predstavlja konkretnu reč ili frazu koja se pojavljuje u tekstu i ukazuje na kategoriju aspekta. Obično su iskazi aspekta imenice, ali se mogu pojaviti i u obliku priloga, prideva i glagola, obično kada je reč o implicitnom iskazu aspekta. Svaki formirani klaster iskaza aspekta određenog entiteta  $e_i$  ukazuje na jedinstven aspekt  $a_{ij}$ .

U (Liu, 2012) osnovni **model entiteta** je formalizovan na sledeći način:

*Entitet  $e_i$  je predstavljen kao celina ili određenim skupom atributa  $A_i = \{a_{i1}, a_{i2}, \dots, a_{in}\}$ .  $e_i$  se može iskazati bilo kojim konačnim skupom izraza entiteta  $\{ee_{i1}, ee_{i2}, \dots, ee_{is}\}$ . Svaki aspekt  $a_{ij} \in A_i$  entiteta  $e_i$  se može iskazati bilo kojim konačnim skupom iskaza aspekta  $\{ae_{ij1}, ae_{ij2}, \dots, ae_{ijm}\}$ .*

3. Ekstrahovanje nosilaca mišljenja iz tekstova ili strukturiranih podataka i njihovo kategorisanje. Zadatak je analogan prethodnim zadacima. Nosilac mišljenja može biti pojedinac ili organizacija. U recenzijama i blogovima nosilac je obično autor objave. Nosioci mišljenja su značajniji za novinske članke jer je u njima obično eksplicitno iskazan autor ili organizacija koja zastupa mišljenje. Međutim, u određenim uslovima identifikovanje nosioca mišljenja u sadržajima sa društvenih medija može biti značajno, npr. identifikovanje i razlikovanje reklame i pojedinca koji citira reklamu.
4. Ekstrahovanje vremena kada su mišljenja iskazana i standardizovanje različitih formata.
5. Klasifikacija sentimenta prema aspektima podrazumeva utvrđivanje polariteta sentimenta u iskazanom mišljenju o određenom aspektu  $a_{ij}$  ili dodeljivanje numeričke ocene sentimenta.
6. Generisanje petorke mišljenja  $(e_i, a_{ij}, s_{ijk}, h_k, t_l)$  iskazanih u dokumentu  $d$  na osnovu prethodnih zadataka. Ovo je zadatak koji se u praksi prilično teško realizuje.

Na osnovu izloženog autor rada (Liu, 2012) definisao je **model dokumenta u kojem je iskazano mišljenje** na sledeći način:

*Dokument u kojem je iskazano mišljenje  $d$  sadrži mišljenja skupa nosilaca mišljenja  $\{h_1, h_2, \dots, h_p\}$  o skupu entiteta  $\{e_1, e_2, \dots, e_j\}$  i podskupu njihovih aspekata iskazanih u određenom vremenskom periodu.*

Sentiment analiza ili *opinion mining* zasnovan na gore navedenom radnom okviru naziva se i aspektno orijentisana sentiment analiza (ili aspektno orijentisan *opinion mining*).

U osnovi sentiment analize prevashodno se nalaze nadgledane tehnike klasifikacije. One zahtevaju unapred pripremljene skupove za obučavanje koji obuhvataju jasno razgraničene i označene primere svake klase. Tekstualni podaci prikupljeni sa sajtova društvenih medija moraju se adekvatno pripremiti kako bi se skup prezentovao odabranim metodama i tehnikama *data mining*-a, što se obično realizuje dodavanjem metapodataka skupu. Svaka oznaka metapodatka upotrebljena da označi element skupa podataka naziva se **anotacijom** inputa. Sprovedene anotacije moraju biti tačne i relevantne zadatku da bi se postiglo efikasno obučavanje algoritma. Iz ovog razloga anotacija tekstova predstavlja kritičnu vezu u razvoju aplikacija za obradu prirodnog jezika (Pustejovsky, et al., 2012). Skupovi tekstova se nazivaju **korpusom**, a jedan skup tekstova anotiran na osnovu iste specifikacije se naziva **anotiranim korpusom**. Anotirani korpusi se koriste za obučavanje algoritama mašinskog učenja.



## 2. SENTIMENT ANALIZA

Sentiment analiza predstavlja proces klasifikacije koji se obavlja na jednom od tri moguća nivoa. Sentiment klasifikacija *na nivou dokumenta* ima za cilj da klasifikuje dokumente na one koji iskazuju pozitivan, odnosno negativan sentiment. Uzima u obzir celokupni dokument i polazi od premise da se u dokumentu raspravlja o jednoj temi. Ovaj vid klasifikacije često ne pruža dovoljno detaljnosti o preovlađujućem mišljenju potrošača o različitim aspektima posmatranog entiteta, što je zahtev mnogih aplikacija (Medhat, et al., 2014). Za viši stepen detaljnosti radi se aspektno orijentisana analiza koja se bavi klasifikacijom sentimenta na nivou rečenice ili fraze prema različitim aspektima analiziranog entiteta. Sentiment analiza *na nivou rečenice* zahteva da se prvobitno razdvoje objektivne od subjektivnih rečenica, potom se sentiment analizom utvrđuje da li je iskazan pozitivan ili negativan stav u subjektivnim rečenicama. Treći nivo sentiment analize podrazumeva klasifikaciju *reči ili fraza* prema polaritetu iskazanog sentimenta.

S obzirom na porast e-trgovine većina istraživanja iz domena sentiment analize je fokusirana na ekonomska i tržišna istraživanja. Bave se pitanjem da li polaritet i/ili obim onlajn recenzija imaju merljiv i značajan uticaj na odluke o kupovini. Potrebno je dobro shvatiti ekonomski uticaj onlajn recenzija na poslovanje jer rezultati sentiment analize recenzija imaju značajne implikacije na stepen truda koji bi kompanije trebalo da ulože u praćenje i upravljanje svojom onlajn reputacijom. Jedan od načina za sticanje dobre reputacije jeste dobijanje mnoštva pozitivnih onlajn recenzija o proizvođaču/prodavcu ili konkretnom proizvodu koji nude. Za mnoga poslovanja mišljenja iskazana na društvenim medijima pretvorila su se u određenu vrstu virtuelne valute koja može da popularizuje ili uništi proizvod na tržištu (Balahur, 2011). Brojna istraživanja dokazuju značajan ekonomski efekat sentiment polariteta u onlajn recenzijama na poslovanje kompanije i odluke kupaca ((Archak, et al., 2007), (Basuroy, et al., 2003), (Berger, et al., 2010), (Chevalier, et al., 2006), (Dellarocas, et al., 2007), (Liu, et al., 2007), (Melnik, et al., 2005), (Pavlou, et al., 2006), (Zhu, et al., 2006)).

Način na koji potrošači iskazuju svoja mišljenja na društvenim medijima olakšava kompanijama da prikupe podatke od interesa. Njihovom analizom ne samo da se olakšavaju brojni zadaci koji su zahtevali značajne napore marketing departmana, poput jednostavnijeg istraživanja tržišta za potrebe oglašavanja i poslovne inteligencije, nego se mogu detektovati i generalne promene na tržištu, otkriti nove tehnologije, identifikovati tržišta na kojima su potrebni određeni proizvodi. Dominantan izvor podataka za sentiment analizu predstavljaju recenzije proizvoda. Izvori recenzija su obično sajtovi za recenziranje i forumi, ali mogu biti i blogovi, sajtovi društvenih mreža, *Twitter*, i drugi društveni mediji.

Istraživanja iz domena sentiment analize i uspešni primeri iz prakse prevashodno se odnose na oblast analize kritika pri odabiru proizvoda, upravljanje reputacijom i odnosima sa javnošću kroz istraživanje javnog mnjenja. Međutim, brojne aplikacije sentiment analize su razvijene za predviđanje trenda prodaje ili tržišta hartija od vrednosti, analizu sentimenta u vestima, za

praćenje i predviđanje izbora, političkih tema, praćenje neprijateljske i negativne komunikacije u političkom svetu, kao i u zdravstvu. Zavisno od oblasti primene istraživanja iz domena sentiment analize usmerena su ka unapređenju različitih socijalnih, ekonomskih, političkih i psiholoških aspekata svakodnevnog života (Balahur, 2011). Saznanja i zaključci izvedeni iz analize recenzija jednog domena često nisu primenjivi u drugom domenu, npr. generalno posmatrano, onlajn recenzije imaju snažniji uticaj na kupovinu skupljih proizvoda ili usluga, dok značajno manje utiču na prodaju jeftinijih (Pang, et al., 2008). Autori u radu (Pang, et al., 2008) takođe ističu da su konfliktni nalazi zabeleženi i unutar istog domena. Različiti segmenti potrošača mogu drugačije reagovati, na primer, potrošači sa snažnijom motivacijom za kupovinom mogu shvatiti onlajn recenzije ozbiljnije od drugih. Dodatno, pojedina istraživanja pokazuju da pozitivne recenzije imaju uticaja, dok druga dokazuju uticaj negativnih recenzija. Autori rada (Archak, et al., 2007) otkrili su da fraze poput „fënomenalan foto-aparat“, „odlična kamera“ i slično imaju negativan uticaj na tražnju. Ukoliko recenzija pored ovog teksta ne sadrži i druge detalje, potrošači doživljavaju ovakve fraze kao hiperbole i tretiraju datu recenziju kao nepouzdanu. Usled činjenice da brojne onlajn recenzije sadrže hiperboličan stil pisanja, autori su ustanovili da fraze poput „pristojan kvalitet“, „dobro pakovanje“ i slično imaju negativan efekat na potrošačke odluke o kupovini jer rečima poput „pristojno“ pripisuju ravnodušan ton. Navedena saznanja ukazuju na razlike u opštem polaritetu određenih reči i kontekstualnom polaritetu reči ili fraza. Superlativi poput „fënomenalan“ jasno nose pozitivan sentiment bez obzira na to da li su autori recenzije uspeali da uvere čitaoce da je proizvod dobar, odnosno, kontekst je potreban kako bi se obrazložio ekonomski efekat smanjenja prodaje, ne i za interpretaciju same recenzije (Pang, et al., 2008). Brojne studije pokazuju da zapravo broj onlajn recenzija oslikava WoM, pa se u brojnim slučajevima ekonomski uticaj ne ogleda samo u broju pozitivnih recenzija nego u ukupnom broju pozitivnih i negativnih.

Svrha ovog poglavlja jeste da se doprinosi istraživanja koje je predmet doktorske disertacije smeste u širi kontekst. Kako je opus istraživanja izuzetno širok, u narednim poglavljima prikazana su samo reprezentativna i novija dostignuća u relevantnim oblastima primene sentiment analize.

## **2.1. Sentiment analiza i mišljenja o proizvodima i uslugama**

Uspeh i proliferacija veb 2.0 aplikacija učinili su da društveni mediji i korisnički generisan sadržaj postanu značajna prednost aplikacija koje podržavaju ključnu poslovnu inteligenciju i marketing (Yang, et al., 2015/2). Sajtovi društvenih medija omogućavaju korisnicima slobodnu razmenu mišljenja i iskustava o procesu kupovine, kvalitetu proizvoda i uslugama. Zbog svog značaja i široke rasprostranjenosti, onlajn recenzije postaju ključni tip korisnički generisanog sadržaja i zauzimaju centralno mesto u istraživanjima iz domena sentiment analize.

Za kompanije recenzije korisnika predstavljaju odličan izvor podataka za sticanje kompetitivne inteligencije korisne za razvoj i unapređenje marketinga. Kroz korisničke recenzije preduzeća dobijaju iskrene, nepristrasne povratne informacije o svojim proizvodima, kao i o proizvodima konkurencije. Na ovaj način se mogu identifikovati tržišni segmenti kojima proizvod najviše odgovara, sagledati razlike između proizvoda koji preduzeće nudi i preferencija potrošača, kao i razlike u mogućnostima preduzeća i konkurencije. Interpretacijom velike količine podataka i pridruženih osećanja kompanije mogu unaprediti poslovno odlučivanje identifikovanjem novih ideja i rešenja za njihove tehnološke ili ekonomske probleme (Balahur, 2011).

Prvi pokušaji rađeni su u domenu filmske industrije, odnosno nad recenzijama filmova (Pang, et al., 2002), nad recenzijama tehničkih proizvoda, knjiga, alternativne muzike i filmova (Dave, et al., 2003), automobilske i bankarske industrije i turističkih destinacija (Turney, 2002). Klasifikacija je vršena na grubom nivou granulacije – na nivou dokumenta. Kategorizacija orijentacije sentimenta iskazanog u dokumentima (recenzijama) vršena je na osnovu pridruženih indikatora sentimenta, poput broja zvezdica, ili na osnovu prisustva pozitivnih i negativnih prideva i priloga. Centralni koncept je baziran na automatskom učenju modela sentiment klasifikacije upotrebom skupa prethodno kategorizovanih dokumenata kao podataka za obučavanje, te se dodatno označavanje prikupljenih dokumenata nije zahtevalo. Razvijeni model se koristi da klasifikuje ili predvidi sentiment kategoriju (pozitivnu, negativnu i, ređe, neutralnu) kojoj novi dokument pripada. Autori u radu (Dave, et al., 2003) opisuju idealan *opinion-mining* alat koji bi obradio rezultate pretrage, generisao listu atributa koji karakterišu kvalitetan proizvod, te agregirao loša, dobra i mešovita mišljenja o proizvodima. Veliki deo istraživanja sprovedenih nakon ovog rada izvode iste zaključke, pre svega da *opinion mining* alati treba da ekstrahuju i analiziraju mišljenja o različitim aspektima proizvoda/usluga.

Rapidni porast broja korisničkih recenzija otežava kompanijama i potrošačima sticanje sveobuhvatne slike o preovlađujućem mišljenju potrošača. Otuda se polako napušta tradicionalna sentiment analiza na nivou dokumenta i sve veći broj istraživanja ispituje mogućnost detaljnije analize mišljenja i osećanja: na nivou reči/fraze ((Breck, et al., 2007), (Zirn, et al., 2011)), na nivou rečenice ili paragrafa ((Gamon, et al., 2005), (Kim, et al., 2006/1), (Wiegand, et al., 2009)) ili na novou dela rečenice ((Wilson, et al., 2005/2), (Choi, et al., 2008)). Dodatno, većina recenzenata iskazuje pozitivan i negativan sentiment u jednoj recenziji i dotiče veći broj karakteristika/aspekata proizvoda (Broß, 2013), što je dovelo do razvoja aspektno orijentisane sentiment analize, koja se bavi analizom sentimenta potrošača prema pojedinačnim aspektima proizvoda. Na ovaj način se obezbeđuje detaljnija analiza ciljnog proizvoda prema njegovim karakteristikama. Za realizaciju detaljnije analize osećanja neophodno je kreirati reprezentativne korpuse pridruživanjem specifično dizajniranih oznaka recenzijama potrošača. Ovaj postupak se naziva anotiranjem korpusa.

U radu (Hu, et al., 2004) autori su sproveli sumiranje kritika potrošača na osnovu karakteristika ocenjenog proizvoda. Njihov korpus čini 113 korisničkih recenzija tehničkih proizvoda (digitalne kamere, DVD plejeri, mp3 plejeri i mobilni telefoni), 4555 rečenica i 81855 tokena. U

tradicionalnim pristupima sumiranje se vršilo odabirom podskupa originalnih rečenica iz recenzija kako bi se oslikalo sveukupno mišljenje potrošača. Pristup koji su autori zauzeli razlikuje se od tradicionalnog, a realizuje se u tri koraka. Koristeći tehnike *data mining*-a i NLP-a prvo se vrši analiza samo onih karakteristika proizvoda o kojima su potrošači iskazali mišljenje. Autori su identifikovali 69 karakteristika proizvoda, a u postupku pripreme korpusa posebno su označavali njihovo implicitno pominjanje. Zatim se identifikuju rečenice u kojima je iskazano mišljenje i utvrđuje se njihov sentiment polaritet. Ove rečenice moraju obuhvatati jednu ili više karakteristika identifikovanih u prethodnom koraku. Kako bi se utvrdila orijentacija iskazanog sentimenta, koriste se metodi NLP-a za identifikovanje skupa prideva kojima se obično iskazuju mišljenja, a za svaku identifikovanu reč se utvrđuje semantička orijentacija (pozitivna ili negativna) korišćenjem *bootstrapping* tehnike<sup>19</sup> i *WordNet*-a<sup>20</sup>. U poslednjem koraku vrši se sumiranje rezultata na strukturiran način koji omogućava potencijalnim kupcima da sa lakoćom sagledaju mišljenja postojećih potrošača o foto-aparatima u celini, kao i da praćenjem odgovarajućih linkova pregledaju pojedinačne karakteristike proizvoda i njegove pozitivne i negativne aspekte prema ocenama ostalih korisnika. Proizvođačima se daje mogućnost da kombinuju izveštaje sa različitih maloprodajnih sajtova kako bi izgenerisali jedinstveni izveštaj za svaki svoj proizvod. Autori su takođe ukazali na efikasnost upotrebe asocijativnih pravila u fazi pripreme podataka kako bi se identifikovale frekventne reči (imenice) ili karakteristike proizvoda, dok u radu (Yang, et al., 2015/1) ističu da se orijentacija obližnjih prideva može utvrditi upotrebom informacija o sinonimima i antonimima u *WordNet*-u.

Autori rada (Kim, et al., 2014) su se bavili klasifikacijom sentimenta prema aspektima proizvoda. Njihova metodologija omogućava identifikovanje pozitivnog i negativnog mišljenja o mobilnim telefonima korišćenjem reči kamera, zvuk, tastatura i baterija kao attribute, a Ekran, Mreža, AP (procesor), Veličina, Kamera i Audio kao šest predefinisanih klasa. Za razliku od svojih prethodnika, autori su postigli detaljno merenje iskazanog mišljenja, koje omogućava da se utvrdi informacija relevantna za konkretni uređaj, kao i deo specifikacije mobilnog telefona koji čini njegovu konkurentsku prednost u odnosu na druge proizvode. Autori su koristili domenski rečnik sačinjen od 500 reči kojima su pridružene težine od -5 do +5 za svaku sentiment reč, što se pokazalo kao poteškoća u primeni implementiranog pristupa u drugim domenima.

U radu (Yang, et al., 2015/2) autori se bave problemom domenske zavisnosti. Predlažu tehniku za generisanje domensko-nezavisnog bazičnog klasifikatora koja se oslanja na mehanizam koobučavanja radi prilagođavanja bazičnog klasifikatora specifičnom domenu primene. Empirijska evaluacija rezultata pokazuje da je predložena tehnika superiorna u odnosu na postojeće pristupe iz dva razloga. Podaci za obučavanje iz izvornog domena mogu se pripremiti upotrebom rečnika opšte namene i nekoliko bazičnih pozitivnih i negativnih reči koje karakterišu mišljenja, čime se štede vremenski i ljudski resursi. Pošto se koristi domen-nezavisan bazni klasifikator, problem heterogenosti između izvornog i ciljnog domena se može eliminisati.

<sup>19</sup> Detaljnije objašnjenje tehnike dostupno na: <http://www.stat.rutgers.edu/home/mxie/rcpapers/bootstrap.pdf>

<sup>20</sup> Dostupan na <https://wordnet.princeton.edu/>

Autor rada (Broß, 2013) bavio se aspektno orijentisanom sentiment analizom, koja podrazumeva identifikovanje aspekta proizvoda o kojem je pisana recenzija i klasifikovanje sadržaja na pozitivan ili negativan. Za potrebe istraživanja razvio je korpus sačinjen iz 310 recenzija hotela i korpus od 370 recenzija digitalnih foto-aparata. Označavanje sadržaja vršeno je na nivou rečenice. Svakoju rečenici su pridružene informacije o: funkciji diskursa poput saveta, poređenja i dr., sentiment polaritetu, temi – definisane su 22 teme za hotele i 31 tema za foto-aparate, i polarizovanim činjenicama. Takođe, označeni su entiteti u fokusu, da li se rečenica odnosi na nerealni događaj i stepen pouzdanosti načinjene anotacije. Ukoliko reč u rečenici nosi sentiment izvršeno je njeno označavanje, uz orijentaciju sentimenta i intenzitet, cilj sentimenta i promenu sentimenta. U klasifikaciji sentimenta primenjen je pristup zasnovan na leksikonima i tehnike nadgledanog učenja. Takođe, autor se bavio ispitivanjem mogućnosti automatskog generisanja sentiment rečnika prema predefinisanim aspektima proizvoda, kao i rečnika svih sentiment pojmova za date domene.

Realizacija aspektno orijentisane sentiment analize često zahteva naporan i vremenski zahtevan posao stvaranja odgovarajućih resursa, odnosno označavanja obučavajućeg skupa (korpora) za metode i tehnike mašinskog učenja. Otuda su u nastavku prikazani neki od reprezentativnih korpusa u istraživanjima iz oblasti analize sentimenta iskazanog u onlajn sadržajima.

Jedan od kompleksnijih korpusa predstavlja 2010 JDPa Sentiment Corpus (Kessler, et al., 2010), kojeg čini 335 poruka preuzetih sa bloga o automobilima, sačinjenih iz 13126 rečenica i 223001 tokena<sup>21</sup>. Autori su odvojeno označavali eksplicitno i apstraktno pominjanje sentiment ciljeva, specifičnih tipova entiteta (osoba, organizacija, lokacija i vreme), relacije između pominjanja (poput: deo, instanca, karakteristika), sentiment reči i orijentacije sentimenta, reči koje menjaju sentiment (engl. *sentiment shifter*), negacije, intenzifikatora i neutralizatora sentimenta.

U radu (Toprak, et al., 2010) prezentovan je ručno označeni korpus sačinjen iz recenzija koje su usmerene na dva različita domena: onlajn univerziteti i onlajn usluge. Skupom je obuhvaćeno 240 generalnih recenzija univerziteta (2786 rečenica) i 234 recenzija usluga (6091 rečenica). Označavanje sadržaja je izvršeno na nivou rečenice i na nivou reči. U rečenicama u kojima je identifikovano prisustvo eksplicitnog mišljenja ili polarizovane činjenice izvršeno je njihovo označavanje, uz označavanje orijentacije iskazanog sentimenta. Na nivou reči označen je cilj sentimenta, izraz, promena sentimenta, intenzifikatori i negacija.

Kako se fokus istraživanja u domenu sentiment analize pomera sa klasifikacije celokupnih dokumenata prema polaritetu iskazanog sentimenta na identifikovanje sentimenta prema aspektima i temama o kojima se piše, automatizacija procesa ekstrahovanja eksplicitnih i implicitnih aspekata iz recenzija predstavlja noviji pravac istraživanja. Jedno od prvih istraživanja rađenih na ovu temu predstavljeno je u radu (Liu, et al., 2005). Autori su predložili sekvencijalni metod učenja baziran na jezičkim obrascima, kojim je omogućeno korisnicima da

<sup>21</sup> Token predstavlja jednu reč u tekstu.

sagledaju prednosti i mane svake karakteristike proizvoda (primer digitalnih foto-aparata). Na ovaj način se potencijalnim kupcima daje paralelni prikaz mišljenja ostalih korisnika o pojedinačnim karakteristikama proizvoda, čime im je olakšano donošenje odluke o proizvodu koji će kupiti, dok proizvođačima olakšava sticanje znanja iz domena marketinga, merenje uspešnosti i poređenje kvaliteta proizvoda sa konkurentskim proizvodima. Autor u radu (Shen, 2015) opisuje metod za automatsko ekstrahovanje teme i njegovog sentimenta iz recenzija hotela. Pristup polazi od predefinisane seta eksplicitnih tema na osnovu kojih se ekstrahuju nepoznate, implicitne teme. Prethodna istraživanja su fokusirana na polaritet iskazanog sentimenta, dok se u istraživanju prezentovanom u ovom radu kalkuliše jačina sentimenta na osnovu leksikona. Prednost prezentovanog modela se ogleda u kombinovanju domenskog rečnika sa opštim sentiment rečnikom, upotrebi *bootsrap* algoritama za ekstrahovanje implicitnih tema iz recenzija (na osnovu eksplicitnih) i linearne regresije za kalkulisanje jačine iskazanog sentimenta. Dosadašnja istraživanja su se bavila eksplicitno iskazanim aspektima i temama, te je ovo istraživanje utoliko značajnije jer predstavlja jedan od prvih pokušaja istraživanja implicitno iskazanih tema.

Sa ciljem da pomognu organizacijama u evaluaciji mišljenja i ponašanja klijenata prema kompaniji i konkretnom proizvodu, autori rada (Dalal, et al., 2014) razvili su sentiment klasifikator baziran na *fuzzy* funkcijama, kojima oponašaju dejstvo različitih lingvističkih elemenata, poput modifikatora. Klasifikator vrši binarno i detaljnije klasifikovanje sentimenta iskazano u recenzijama i komentarima korisnika sa *Twitter*-a. Za evaluaciju procedure koristili su skup od 3000 korisnički generisanih recenzija o proizvodima (tableti, e-čitači knjiga, telefoni i laptop računari) i ocene recenzija u vidu 1-5 zvezdica. Dobijeni rezultati su pokazali da je prezentovani pristup zasnovan na lingvističkim komponentama efikasniji sa porastom granularnosti zadatka u odnosu na druge aplikacije.

U radu (Yang, et al., 2015/1) predstavljen je FOBPRM (engl. *Feature Ontology Based Product Review Miner*) sistem koji poluautomatski razvija stablo ontologija semantičkog prostora reči „telefon” i ekstrahuje iz recenzija najreprezentativnije izraze i mišljenja potrošača o telefonima, koji predstavljaju parove atribut-sentiment. Sistem kalkuliše polaritet ovih parova na osnovu entropije i generiše sveukupni izveštaj za potrošače i proizvođače.

S aspekta potrošača, korisničke recenzije daju korisne informacije koje skraćuju vreme pretrage i rezultuju efikasnijim odlukama o kupovini (Yang, et al., 2015/2). Više od 80% potrošača očekuje na sajtu prodavca implementiranu mogućnost konsultovanja drugih korisnika ili profesionalaca pre samog čina kupovine (D'Avanzo, et al., 2013). Otuda je razvoj aplikacija sentiment analize za potrebe e-trgovine široko rasprostranjena oblast istraživanja. Sajtovi na kojima se obavlja elektronska trgovina eksploatišu mogućnosti korisničkih recenzija, analiziraju, klasifikuju i sumiraju mišljenja iz recenzija o proizvodima koje posetioci mogu koristiti za poređenja, savetovanje i preporuku. Autori u radu (D'Avanzo, et al., 2015) uvode kognitivno zasnovanu proceduru kojom analiziraju mišljenja korisnika sa specifičnih tipova tržišta i vizuelno ih sumiraju kako bi se ublažila preopterećenost kupaca velikom količinom recenzija i ubrzala njihova

kupovna aktivnost. Njihov pristup oponaša zonu narednog razvoja Vigotskog<sup>22</sup> koja je široko eksploatisana u zajednicama kolaborativnog učenja.

Sistemi za davanje preporuka na osnovu korisničkih recenzija predstavljaju tipične aplikacije inkorporirane u sajtove e-trgovine. Zahvaljujući ovim sistemima korisnici sajta ne pretražuju direktno recenzije, već dobijaju preporuku proizvoda koja najviše odgovara njihovim preferencijama. Međutim, ovakvi sistemi se uglavnom oslanjaju na strukturirane metapodatke pri davanju preporuka, poput pridruženih ocena u vidu zvezdica, i ignorišu tekstualni sadržaj kao najvažniji izvor informacija u recenzijama. Autori u radu (Ganu, et al., 2009) smatraju da bi se korisničko iskustvo značajno unapredilo ukoliko bi se u obzir uzela i struktura sadržaja recenzija, odnosno ukoliko bi se identifikovali delovi recenzija koji su usmereni ka određenim karakteristikama proizvoda (u njihovom slučaju, hrana, ambijent, cena i usluga), kao i sentiment recenzenta ka svakoj karakteristici (pozitivan, negativan ili neutralan). U analizi recenzija upotrebili su kombinaciju ovih informacija i metapodataka povezanih sa proizvodom. Autori koriste saznanja o karakteristikama proizvoda i sentimentu za unapređenje korisničkog iskustva tokom pristupa i čitanja recenzija i fokusirani su na unapređenje tačnosti preporuka restorana na osnovu recenzija. Cilj istraživanja je da se obezbedi bolje razumevanje obrazaca korisničkih recenzija i razvoj alata za efikasniju pretragu, razumevanje i pristup korisničkim recenzijama. Pristup se bazira na kombinovanju NLP, mašinskog učenja i kolaborativnog filtriranja kako bi se iskoristilo bogatstvo detaljnih informacija iz recenzija. Rezultati analiza dobijenih primenom razvijenih tehnika za klasifikaciju i analizu recenzija upotrebljeni su za unapređenje personalizovanih preporuka veb korisnicima. Rezultati istraživanja pokazuju da upotreba detaljnih tekstualnih informacija rezultuje kvalitetnijim opštim i personalizovanim predviđanjem ocena recenzija u odnosu na upotrebu numeričkih ocena koje daju korisnici.

## **2.2. Sentiment analiza u finansijskom sektoru**

U opisu ponašanja investitora bihevioralne finansije se oslanjaju na psihološke koncepte, tvrdeći da proces odlučivanja nije uvek vođen racionalnim kalkulacijama, nego je često pod uticajem emocija i drugih psiholoških faktora (De Bondt, et al., 1994). Informativni mediji predstavljaju jedan od faktora koji utiče na percepciju investitora (stavovi o budućem toku gotovine i diskontne stope, koji nisu pod uticajem makroekonomskih faktora ili faktora specifičnih za kompaniju, nego su povezani sa emocionalnim rezonovanjem). Vesti utiču na percepciju investitora i mogu čak delovati kao primarni izvor nestabilnosti na finansijskim tržištima ukoliko je stil njihovog prezentovanja u medijima, ili sam sadržaj, obogaćen elementima emocije i neizvesnosti (Van de Kauter, et al., 2015). Autori rada (Devitt, et al., 2007) navode da su fluktuacije javnog mišljenja o kompaniji povezane sa fluktuacijama njihovih cena akcija. Sve

---

<sup>22</sup> Rastojanje između najtežeg zadatka koji neko može samostalno da uradi i najtežeg zadatka koji može da uradi uz nečiju pomoć.

ovo je aktiviralo interesovanje za primenom *opinion mining*-a i sentiment analize u finansijskom sektoru i domenu planiranja i unapređenja prodaje. *Opinion mining* i sentiment analiza se uspešno koriste za praćenje mišljenja javnosti kroz vreme. Dobijena saznanja se koriste za potrebe finansijskih i tržišnih analiza, u predviđanju kriznih situacija i za alarmiranje.

Autori rada (Van de Kauter, et al., 2015) ističu da vesti asimetrično utiču na tržište hartija od vrednosti: negativne vesti imaju jači uticaj na nestabilnost nego pozitivne vesti i mogu imati različite efekte na prihode, zavisno od celokupne tržišne situacije. U svom radu autori su ispitali upotrebu sentiment analize za detektovanje pozitivnog i negativnog sentimenta iskazanog o određenim kompanijama u finansijskim novinskim člancima. Primena sentiment analize nad finansijskim člancima omogućava istraživačima iz domena finansija da identifikuju pozitivne i negativne vesti o kompaniji na automatizovan način i posledično da obrade veću količinu podataka za kraće vreme, što može voditi ka novim saznanjima o uticaju vesti i medija na tržište hartija od vrednosti.

U radu (Bollen, et al., 2011) autori su upotreбили analizu *Granger*-ove kauzalnosti i samoorganizujuće *fuzzy* neuronske mreže kako bi ispitali hipotezu da javno raspoloženje mereno alatima za praćenje raspoloženja može da se upotrebí za predviđanje promena završnih vrednosti industrijskog indeksa *Dow Jones (DJIA)*. Rezultati su pokazali da se tačnost predviđanja *DJIA* može značajno poboljšati uključivanjem specifične dimenzije javnog raspoloženja (tačnost od 87.6% u predviđanju promena u *DJIA* vrednostima i smanjenje srednje apsolutne greške izražene u procentima (MAPE – *mean average percentage error*) za više od 6%).

Istraživanja pokazuju da su finansijske informacije predstavljene u novinskim člancima i vestima u snažnoj korelaciji sa društvenim fenomenima o kojima se iskazuje mišljenje na blogovima i forumima. Radovi (Bar-Haim, et al., 2011) i (Feldman, et al., 2011) bave se identifikovanjem eksperata iz domena hartija od vrednosti na mikroblogovima (*Twitter*-u) i primenom sentiment analize na sadržaje koje objavljuju. Istraživanje autora je pokazalo da se primenom sentiment analize nad sadržajima eksperata dobija relativno visoka tačnost predviđanja rasta akcija.

Autori u radu (Liu, et al., 2007) izučavaju problem rudarenja sentimenta u informacijama sa blogova i ispituju načine kojima se takve informacije mogu koristiti za predviđanje prodaje proizvoda. Oni su predložili ARSA model, skraćeno od engleskih reči *autoregressive sentiment-aware model* (autoregresivni model svestan osećanja), koji odražava efekte osećanja i proteklih prodajnih učinaka na buduće performanse prodaje. Opsežna eksperimentisanja nad skupom podataka o filmovima upoređena su sa alternativnim modelima koji nisu uzeli u obzir informacije o osećanjima. Rezultati ovih poređenja pokazuju da je predloženi pristup efikasniji i superiorniji.

Pored navedenog, rađena su brojna istraživanja o primeni sentiment analize u različitim ekonomskim i finansijskim aplikacijama poput predviđanja tržišta (u radu (Khadjeh Nassirtoussia, et al., 2014) dat je i detaljan pregled istraživanja u ovoj oblasti, (Khadjeh



Nassirtoussia, et al., 2015), (Sprenger, et al., 2014)), utvrđivanja sentimenta finansijskih blogera prema kompanijama i njihovim akcijama (O'Hare, et al., 2009). Finansijski korpusi za potrebe obučavanja i evaluacije rada algoritama mašinskog učenja razvijeni su manuelnom anotacijom ((O'Hare, et al., 2009), (Mostafa, 2013), (Ghiassia, et al., 2013)) i automatski na bazi promena u cenama akcija (Koppel, et al., 2006).

### **2.3. Primena sentiment analize u poslovnoj inteligenciji**

U turbulentnom poslovnom okruženju fokusiranom na potrošače, uspeh kompanije umnogome zavisi od toga koliko brzo i na koji način one reaguju na promenljive tržišne uslove i kako u takvim okolnostima stvaraju vrednosti od interesa za svoje potrošače. Upotrebom podataka, tehnologija i analiza poslovna inteligencija pruža kompanijama uvide i saznanja neophodna za donošenje odluka i sprovođenje akcija sa ciljem ostvarivanja ekonomske vrednosti (rast preduzeća, smanjenje troškova, privlačenje novih i zadržavanje postojećih potrošača) (Herschel, 2011). Kao takva, poslovna inteligencija je usko povezana sa oblastima primene sentiment analize koje su opisane u prethodnim poglavljima.

Sistemi poslovne inteligencije prikupljaju podatke iz svih delova organizacije, omogućavaju njihovu analizu i pripremu personalizovanih izveštaja za svakog pojedinca u organizaciji. Međutim, da bi se poslovna inteligencija dalje razvila, potrebno je izmeniti način na koji se ona koristi. Usled činjenice da društveni mediji predstavljaju mesto na kojem se danas nalaze potrošači, neophodno je da ih kompanije uključe kao jedan od izvora podataka za sisteme poslovne inteligencije, te da se i na taj način vrši monitoring preferencija, namera i interesa potrošača, kao i analiza konverzacija i sadržaja razmenjenih na društvenim medijima (Grljević, et al., 2012). Obezbeđenjem smislene i značajne poslovne inteligencije kompanije mogu identifikovati uticajne korisnike na mreži i perspektivne kontakte, mogu pratiti sopstvenu reputaciju na webu ili reputaciju konkretnog proizvoda. U nastavku poglavlja prikazana su istraživanja koja iz različitih uglova posmatraju mogućnosti integracije sistema poslovne inteligencije, društvenih medija i sentiment analize.

U radu (Hiltbrand, 2010) istaknuto je da se dosadašnja postignuća u oblasti poslovne inteligencije oslanjaju na izveštaje i kontrolne table sa ciljem sumiranja i isporučivanja informacija krajnjim korisnicima. Međutim, različiti sajtovi društvenih medija pružaju specifične funkcionalnosti koje kompanije mogu inkorporirati u svoje sisteme poslovne inteligencije i time ih unaprediti, poput ažuriranja statusa i razvijanja komunikacije o određenoj informaciji, društvenih preporuka, rangiranja, folksonomija i pretraga. Autor je predočio da je u ovom kontekstu poslovnu inteligenciju potrebno posmatrati ne kao kolekciju izveštaja i kontrolnih tabli, nego kao skup pojedinačnih objekata, odnosno informacionih komponenti koje bi bile sačinjene iz specifičnih metrika ili kombinacije metrika, što bi omogućilo korisnicima da se uspešnije fokusiraju na najvažnije informacije. Korisnicima sistema poslovne inteligencije

potrebno je omogućiti sa jedne strane kreiranje zajednice kolega, a sa druge strane kreiranje informacionih objekata koji su relevantni za njihovo poslovanje (na primer: praćenje inventara, porudžbine, nabavke). Za svaki informacioni objekat potrebno je definisati prag prihvatljivosti kako bi se omogućilo da se statusi o njemu generišu i dele sa korisnicima sistema samo kada su predefinisani uslovi ispunjeni. Autor sugeriše integraciju društvenog rangiranja i sistemima poslovne inteligencije. Društveno rangiranje bi unapredilo generisanje izveštaja time što bi se oni kreirali kada su postavljeni pragovi prihvatljivosti ispunjeni, nakon čega bi zajednica evaluirala značaj generisanog izveštaja.

Autori rada (Berlanga, et al., 2015) predlažu novu infrastrukturu za poslovnu inteligenciju, baziranu na semantičkim podacima sa ciljem pružanja novih mogućnosti za integraciju tradicionalne i društvene poslovne inteligencije (engl. *social business intelligence*). Predložena infrastruktura prati principe takozvane *Linked Open Data* inicijative, koja podrazumeva takav način objavljivanja podataka koji omogućava da oni budu povezani i obogaćeni kako bi se uspešno pronalazili različiti prikazi istog sadržaja i kako bi se mogle postaviti veze između povezanih resursa. Ovakva infrastruktura podrazumeva povezivanje tradicionalnih korporativnih podataka sa eksternim faktorima koji mogu uticati na poslovanje, poput novih pravnih propisa, finansijskih vesti i podataka sa društvenih medija, kao i automatsko ekstrahovanje sentimenta iz podataka sa društvenih medija, za uspešnije osluškivanje potreba i preferencija potrošača.

Autori u radu (Funk, et al., 2008) predstavljaju MUSING sistem koji primenjuje tehnologije obrade prirodnog jezika, poput ekstrahovanja ontologija, u kontekstu aplikacija poslovne inteligencije. Kako su kvalitativne informacije važne za brojne aplikacije poslovne inteligencije, sistem prikuplja i organizuje mišljenja o poslovnim entitetima (organizacijama, pojedincima ili proizvodima) i vrši predviđanje njihove reputacije. Sistem identifikuje iskaze o reputaciji, pouzdanosti ili kvalitetu, analizira iskazani sentiment i prati ih tokom vremena kako bi kreirao precizniju sliku javnosti o poslovnom entitetu.

## 2.4. Ostale primene sentiment analize

Mišljenja iskazana na mikroblogovima, blogovima, forumima i društvenim mrežama značajno menjaju način na koji zajednice participiraju u elaboraciji zakona i politike, te ova oblast postaje jedna od dominantnih domena primene sentiment analize. Sajtovi društvenih medija koriste se kao izvor podataka u analizi saglasnosti ili nesaglasnosti javnosti sa određenom političkom temom ((Balahur, et al., 2009), (Sobkowicz, et al., 2012)). Sentiment analiza se može primeniti i za potrebe praćenja konzistentnosti i nekonzistentnosti između datih izjava i akcija na državnom nivou, a rezultati izbora se mogu uspešnije predviđati praćenjem sentimenta u diskusijama na blogovima i forumima. Autori u radu (Kim, et al., 2006/1) bavili su se predviđanjem ishoda predsedničkih izbora, dok su se autori rada (Bansal, et al., 2008) bavili ispitivanjem stepena u kom javnost pruža podršku kandidatu i njegovim stavovima.

Primena sentiment analize je široko rasprostranjena i u domenu medicine. Razumevanje iskustva pacijenta tokom zdravstvene zaštite ima centralno mesto u procesu pružanja nege i predstavlja osnovni odraz kvaliteta zdravstvene zaštite. Autori u radu (Greaves, et al., 2013) upotrebili su mašinsko učenje kako bi stekli bolji uvid o nestruktuiranim komentarima pacijenata o zdravstvenoj zaštiti. Komentari su kategorisani u pozitivne i negativne, a sentiment analiza je primenjena sa ciljem predviđanja verovatnoće da će pacijent preporučiti određenu bolnicu, za ocenu čistoće bolnice i ophođenja osoblja prema pacijentima. Autori rada (Hopper, et al., 2015) među prvima se bave mogućnostima upotrebe saznanja dobijenih sentiment analizom o stepenu zadovoljstva pacijenata, pre svega pritužbi, o upravljanju radom bolnice i unapređenju odnosa pacijent-lekar. Metodi sentiment analize su razvijeni i za analizu kliničkih dokumenata koji odražavaju zdravstveno stanje u smislu opservacija lekara i objektivnih informacija poput opisa rezultata pregleda, dijagnoza i intervencija, a sa ciljem procene pozitivnih i negativnih ishoda, procene uticaja trenutnog stanja pacijenta na njegovo buduće zdravlje (Denecke, et al., 2015). Sentiment analiza se uspešno koristi u predviđanju srčanih oboljenja ((Lerner, 2015), (Karystianisa, et al., 2015)), predviđanju poseta hitne pomoći (Blue, 2015) i praćenju epidemije HIV-a (Rivero, 2014). Takođe, intenzivno se primenjuje i u analizi biomedicinskih istraživanja objavljenih u formi naučnih ili novinskih članaka (Swaminatha, et al., 2010).

Kako je istaknuto u radu (Balahur, 2011) sentiment analiza i *opinion mining* unapređuju i pojedine zadatke NLP-a, poput:

- Ekstrahovanja informacija odvajanjem činjenica od mišljenja,
- Utvrđivanja autorskog prava i
- Identifikovanja višeznačnosti reči.

Jedan od najpoznatijih korpusa upotrebljen za unapređenje NLP zadataka, a prevashodno za identifikovanje kontekstualnog polariteta reči i identifikovanje subjektivnosti, je *MPQA* korpus mišljenja<sup>23</sup> (engl. *Multi-Perspective Question Answering*). Ovaj korpus je razvijen 2002. godine i predstavlja najistaknutiji anotirani korpus u oblasti sentiment analize (Kessler, et al., 2010). Korpus čine novinski članci na engleskom jeziku (692 dokumenta sačinjena iz 15802 rečenice) iz raznih izvora, nad kojima je izvršeno ručno označavanje mišljenja i drugih ličnih stavova (verovanje, emocije, osećaji, spekulacije itd.) (Wiebe, et al., 2005). Prvobitno su nad sadržajem označeni lični stavovi na nivou reči/izraza, kojima su dodeljene informacije o intenzitetu i orijentaciji iskazanog sentimenta (Wilson, et al., 2003). U narednom periodu autori su proširivali oznake tako da se obuhvate informacije o direktno iskazanim subjektivnim izrazima i njihovoj kontekstualnoj orijentaciji sentimenta (Wilson, et al., 2005/2), o tome ka kome/čemu je usmeren lični stav i koja je vrsta stava (osećanje, saglasnost, namera, rasprava i spekulacija) (Wilson, et al., 2005/1). Autori rada (Stoyanov, et al., 2008) dodatno su proširili korpus označavanjem informacija o temama iskazanog sentimenta.

<sup>23</sup> Korpus je dostupan na linku [http://mpqa.cs.pitt.edu/corpora/mpqa\\_corpus/](http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/)

## **2.5. Sentiment analiza u visokom obrazovanju**

Sentiment analiza je pronašla široku primenu u pojedinim uslužnim sektorima poput finansijskog, turističkog i zdravstvenog, međutim, malo istraživanja je rađeno na temu primene sentiment analize u visokom obrazovanju. U nastavku poglavlja prikazana su reprezentativna istraživanja iz domena primene sentiment analize u obrazovanju, koja predstavljaju polaznu tačku i motivaciju za istraživanje koje je predmet doktorske disertacije.

Većina dosadašnjih istraživanja u domenu obrazovanja bavila su se mogućnostima unapređenja inteligentnih tutorskih sistema primenom saznanja dobijenih sentiment analizom sadržaja sa društvenih medija. Osnovni razlog njihove brojnosti ogleda se u činjenici da studenti koje obučavaju studenti-tutori postižu bolje rezultate u odnosu na one koje obučava inteligentni tutorski sistem. Istraživanja su pokazala da uključivanje komponente emocija i njihovog predviđanja u softvere može premostiti ovaj jaz (Litman, et al., 2004). Za ove potrebe korpusi su razvijeni ručno. Autori su označavali različite tipove dijaloga studenta i tutora (Nielsen, et al., 2008), a pored toga, vršili su i označavanje neverbalne komunikacije (držanja i gestikulacije) (Grafsgaard, et al., 2013). Neverbalna komunikacija u značajnoj meri signalizira afektivna i motivaciona stanja, a potvrđena je i značajna veza između neverbalne komunikacije i studentskih pitanja, odgovora tutora, direktiva tutora i pozitivnih povratnih informacija od tutora (Grafsgaard, et al., 2013). Studenti koji tokom dijaloga postavljaju obe ruke na lice obično iskazuju nižu produktivnost. Otkrivena saznanja se mogu inkorporirati u inteligentne tutorske sisteme gde bi se određenom gestikulacijom studenta pokretale adekvatne akcije sistema, čime je unapređen način na koji sistem daje odgovore i objašnjenja. Autori rada (Litman, et al., 2004) zauzeli su drugačiji pristup analizi emocionalnih stanja studenata u razgovorima sa tutorima. U pripremi korpusa označavali su negativne, pozitivne i neutralne emocije koje studenti iskazuju u razgovorima, a koje su okarakterisali kao osnovne emocije, dok su kao sporedne emocije definisali slabo iskazane pozitivne, slabo iskazane negativne i mešovite emocije. Dodatno, korpusu su pridružene i oznake domena, da li se iskazana emocija odnosi na nastavne materijale koji su u fokusu istraživanja ili na sam proces podučavanja, kao i oznake emocionalnih reči. Obogaćivanjem korpusa informacijama o domenu i specifičnim emocionalnim rečima može se dublje analizirati emocionalno stanje studenata i na osnovu relacije emocija-domen-reč identifikovati jaz ili problem u podučavanju. Sentiment analizom dijaloga studenta i tutora identifikovni su problemi u podučavanju, stepen u kom studenti shvataju osnovne naučne koncepte, analizirana su pitanja i odgovori, kao i neverbalna komunikacija. Sentiment analiza je potom inkorporirana u inteligentne tutorske sisteme kako bi unapredila podučavanje, davanje odgovora na upite studenata, doprinela efikasnijem razrešenju problema sa kojima se studenti susreću, kao i efikasnijem pokretanju akcija sistema na osnovu posturalnih promena studenta. Kako ovaj vid obučavanja u Srbiji nije aktualizovan, opisana istraživanja predstavljaju dobru polaznu tačku u kontekstu kreiranja neophodnih resursa za sentiment analizu dok u našim uslovima izvori podataka za realizaciju srodnog istraživanja ne postoje ili nisu javno dostupni.

U radovima (Altrabsheh, et al., 2013) i (Altrabsheh, et al., 2014) autori ispituju mogućnosti upotrebe znanja o sentimentu studenata, za podsticanje njihovog angažovanja u nastavi, koje je prevashodno ugroženo usled dosade ili konfuzije. Povratne informacije od studenata o stepenu u kom su shvatili pojedine lekcije, brzini izlaganja profesora i sličnom prikupljane su u stvarnom vremenu. Nad prikupljenim podacima se primenjuje sentiment analiza i rezultati se prikazuju predavaču kako bi bio u mogućnosti da blagovremeno rešava probleme sa kojima se studenti susreću. Rezultati su pokazali da eliminisanje neutralne klase sentimenta u fazi obučavanja algoritama daje bolje rezultate, te da su *Support Vector Machine* i *Complement Naive Bayes* najbolji klasifikatori za dati skup podataka. Razvijeni sistem *Sentiment analysis for Education* koji omogućava preuzimanje povratnih informacija od studenata o predmetu/kursu koji pohađaju u stvarnom vremenu prikazan je u radu (Altrabsheh, et al., 2013). Autori smatraju da se implementacijom ovakvog sistema može značajno pomoći aktuelnim studentima u rešavanju problema sa kojima se suočavaju tokom trajanja nastave i time unaprediti odnos sa studentima. Za realizaciju srodnog istraživanja potrebno bi bilo razviti sistem koji će omogućiti studentima da tokom semestra u elektronskom formatu ostavljaju svoje komentare o kvalitetu nastave. Međutim, efikasnom sentiment analizom podataka sa studentskih foruma o pojedinim predmetima uspešno bi se strukturiralo njihovo mišljenje i identifikovali problemi u nastavi, stilu predavanja ili nastavnom materijalu.

Razumevanje sentimenta studenata skrivenog u porukama postavljenim na forumima može pomoći u identifikovanju studenata koji imaju poteškoće sa predmetom, u optimizaciji pružanja pomoći, identifikovanju pozitivnih i negativnih akcija instruktora ili predavača i u pružanju vrednih povratnih informacija. Autori u radovima ((Wyner, et al., 2009), (Kim, et al., 2010) i (Yoo, et al., 2014)) bavili su se sentiment analizom poruka postavljenih na forumima predmeta iz oblasti informatike. Definisali su set kontekstno nezavisnih činova emocionalnog govora koje studenti univerziteta koriste na forumu kako bi iskazali sigurnost, frustraciju, napetost/tenziju i ljubaznost. U postupku pripreme korpusa prikupljenim sadržajima, koji su dati u formi dijaloga, pridružene su oznake predefinisanih emocionalnih činova. Navedena istraživanja su rađena sa ciljem unapređenja komunikacije između studenta i instruktora, kao i pružanja pomoći studentima koji imaju poteškoće u savladavanju materije. Otkrivena znanja, poput toga da se frustracija češće javlja u dugim diskusijama i da je povezana sa nižim ocenama ili da je visok stepen poverenja u značajnoj korelaciji sa performansama, mogu se iskoristiti u predviđanju performansi studenata, identifikovanju poteškoća na početku semestra, koje će ukazati na različite situacije i vremenske periode kada je studentima potrebno pružiti pomoć.

Autori rada (Wen, et al., 2014) pratili su iskazani sentiment o pojedinim kursovima na forumima, tokom perioda u kom se održavaju, i istražili njegovu vezu sa napuštanjem kursa. U analizama su upotrebljeni podaci o tri kursa koji su održani na *Coursera.org*<sup>24</sup> sajtu za besplatno obučavanje. Ekstrahovane su pozitivne i negativne sentiment reči najviše povezane s analiziranim kursom. Uticaj iskazanog sentimenta na pojedinačnog korisnika analiziran je upotrebom analize

<sup>24</sup> Adresa sajta: <https://www.coursera.org/>.

preživljavanja (engl. *survival analysis*). Ova analiza je omogućila da se ispita način na koji se mišljenje članova foruma izneto tokom jedne nedelje trajanja kursa može upotrebiti za predviđanje njihovog budućeg angažmana na forumima. Rezultati su otkrili isti trend kod sva tri kursa: viši koeficijenti iskazanog sentimenta su povezani sa manjim brojem studenata koji napuštaju kurs. Tokom analize, autori su uspjeli da identifikuju i sentiment reči koje su usko povezane sa kursom, a koje mogu da pomognu u otkrivanju prednosti i problema u njegovoj realizaciji.

Društveni mediji poput *Twitter*-a i *Facebook*-a predstavljaju popularne sajtove među studentima jer pružaju platformu za slobodno iskazivanje pogleda, emocija, stresa i mišljenja o obrazovnom procesu. U istraživanju predstavljenom u radu (Takle, et al., 2015) autori koriste upravo ove izvore za identifikovanje i razumevanje mišljenja i ponašanja studenata na društvenim medijima. Autori su primenili genetičke algoritme i klasifikaciju tekstualnih sadržaja, dok sentiment analiza nije realizovana.

U radu (Sagum, et al., 2015) opisana je upotreba sentiment analize za utvrđivanje stepena pozitivnosti ili negativnosti u komentarima studenata preuzetih iz evaluacionih formulara fakulteta. Cilj istraživanja jeste razvoj aplikacije za automatsku procenu kvaliteta rada profesora na osnovu sentiment analize komentara studenata prikupljenih iz upitnika o proceni fakulteta. *SentiWordNet* je upotrebljen za definisanje stepena pozitivnosti komentara, potom se vrši identifikacija polariteta, označavanje negativije i intenzifikatora, uzimajući u obzir reči u okruženju.

### 3. IMPLEMENTACIJA SENTIMENT ANALIZE

Istraživanja u oblasti sentiment analize, kako je istaknuto u prethodnom poglavlju, prostiru se na brojne domene i sprovode se sa različitim ciljevima. Uprkos razlikama u ciljevima i u konačnim aplikacijama, postupci se mogu grupisati u četiri kategorije zadataka koje nose sa sobom određene izazove (Montoyo, et al., 2012).

1. *Kreiranje resursa za sentiment analizu.* U fokusu istraživanja ove grupe zadataka je kreiranje rečnika i korpusa u kojima se anotiraju izrazi sentimenta prema polaritetu, kao i druge semantičke, sintaksičke ili leksičke informacije. Kao osnovne izazove autori ističu dvosmislenost reči, potrebu za postojanjem resusa za različite jezike, granularnost – mišljenje može biti iskazano rečima, rečenicama ili celim frazama – razlike u načinu iskazavanja mišljenja u različitim tipovima teksta, što nameće potrebu za anotiranjem svakog tipa teksta koji će biti predmet analiza (blogovi, novinski članci, sadržaji sa foruma, sajtova za recenziranje i dr.).
2. *Klasifikacija teksta prema polaritetu iskazanog mišljenja* na pozitivne, negativne i neutralne komentare. Zadatak klasifikacije se može realizovati na nivou dokumenta, rečenice, fraze ili reči. Klasifikacija sentimenta na nivou *dokumenta* polazi od pretpostavke da svaki dokument iskazuje mišljenje o jednoj temi, a osnovni cilj jeste utvrditi da li je u dokumentu iskazana pozitivna ili negativna emocija. Ovaj vid klasifikacije nije primenjiv kada je u dokumentu evaluirano više aspekata ili se porede različiti entiteti (Liu, 2012). Klasifikacija na nivou *rečenice* tretira zasebno svaku rečenicu i polazi od pretpostavke da ona sadrži tačno jedno mišljenje (pozitivno, negativno ili neutralno) (Jagtap, et al., 2013). Klasifikacija na nivou *reči* i *fraze* se realizuje utvrđivanjem polariteta sentimenta reči/fraza koje se koriste za klasifikaciju. Ovi pristupi klasifikaciji se bave otkrivanjem preovlađujućeg sentimenta, ne i aspekta, te ne otkrivaju tačno šta ljudi preferiraju ili ne. Polazna premisa klasifikacije na nivou *aspekta* predstavlja mišljenje sačinjeno iz pozitivnog ili negativnog sentimenta, koje je usmereno ka nekoj osobi, objektu ili aspektu (cilju sentimenta), te je osnovni cilj utvrditi sentiment o entitetu ili aspektu (Liu, 2012). Otuda, klasifikacija na nivou aspekta omogućava detaljniju analizu sentimenta prema aspektima, čime se tačnije identifikuju preferencije ljudi. Kao osnovne izazove klasifikacije teksta, autori rada (Montoyo, et al., 2012) navode potrebu za kontekstualnom analizom iskazanih mišljenja koja zahteva analizu na različitim nivoima (leksičkom, sintaksičkom, semantičkom i pragmatičnom), kao i upotrebu robusnih nadgledanih, polunadgledanih ili nenadgledanih metoda; uvažavanje negacije i njenog delokruga; detektovanje ironije; detektovanje implicitno iskazanog sentimenta u objektivnim rečenicama.
3. *Ekstrakcija mišljenja* se bavi identifikovanjem delova teksta u kojima je mišljenje iskazano, identifikovanjem polariteta iskazanog mišljenja, kao i ka kome/čemu je usmereno. Osnovni

izazovi se odnose na detekciju opsega važenja sentiment iskaza i cilja iskazanog mišljenja (Montoyo, et al., 2012).

4. *Primena sentiment analize* kao samostalnog zadatka, kombinovane sa drugim zadacima obrade prirodnog jezika (sumiranje mišljenja, pronalaženje mišljenja i sl.), ili kao dela kompleksne aplikacije (npr. rangiranje proizvoda prema iskazanom mišljenju korisnika u inteligentnim sistemima za davanje preporuka ili sistemi za detektovanje trendova). Osnovni izazovi se tiču adekvatnog kombinovanja različitih tehnika sentiment analize i metoda iz oblasti u kojoj se primenjuje (Montoyo, et al., 2012).

Istraživanje koje je predmet doktorske disertacije prati navedene zadatke sa ciljem sticanja sveukupne slike o sentimentu studenata. U prvoj fazi istraživanja kreiran je korpus anotiranjem prikupljenih podataka. Na osnovu rezultata anotacije kreirani su potrebni domenski rečnici sentiment pojmova, pri čemu se pod sentiment pojmom podrazumevaju sentiment reči, fraze i izrazi o kojima će biti više reči u [Poglavlju 4.1.3](#). Primenom kreiranih rečnika nad podacima urađena je klasifikacija teksta prema polaritetu iskazanog mišljenja. Anotacije su takođe obezbedile osnovu za sprovođenje aspektno orijentisane sentiment analize sa ciljem ekstrahovanja delova teksta koji su usmereni ka konkretnom cilju i identifikovanja polariteta iskazanog mišljenja. Model koji će automatizovati ovu vrstu analize predstavlja buduće korake istraživanja. U poslednjem koraku istraživanja ilustrovane su mogućnosti primene sentiment analize u poslovanju visokoškolskih institucija. U narednim poglavljima detaljnije su opisani ključni problemi i izazovi sentiment analize, kao i metode i tehnike koje čine njenu osnovu.

### 3.1. Metodologija sentiment analize

Sentiment analiza predstavlja kompleksan analitički proces koji se realizuje kroz nekoliko sekvencijalnih koraka (Deshpande, et al., 2011). Prvi korak je pronalaženje i prikupljanje podataka, zatim sledi priprema podataka, treći korak je ekstrakcija atributa, zatim ekstrakcija sentimenta, klasifikacija sentimenta i njihovo sumiranje, i izveštavanje i vizualizacija rezultata. Istraživanje koje je predmet doktorske disertacije prati prezentovanu metodologiju.

*Pretraga, preuzimanje i čišćenje podataka.* Komentari o proizvodima i uslugama se mogu pronaći na zvaničnim sajtovima kompanija, različitim forumima na internetu, specijalizovanim blogovima, *Facebook* stranicama ili drugim sajtovima društvenih medija. Radi sticanja sveobuhvatnog uvida u diskusije vođene o proizvodima ili uslugama, potrebno je preuzeti sa interneta relevantne komentare. Za ove potrebe se koriste pretraživači veba, tzv. *web crawlers*, koji pretražuju internet, preuzimaju sadržaj ciljane veb stranice i čuvaju ga na lokalnom disku. Alternativno, koristi se kastomiziran programski kod kreiran prema strukturi sajta koji će ciljati i preuzimati sadržaj između određenih HTML tagova.



*Preprocesiranje podataka.* Preuzete veb stranice mogu sadržati komentare o proizvodima ili uslugama koje nisu predmet analize ili su van okvira domena, te je potrebno izvršiti filtriranje teksta kako bi se zadržao sadržaj koji se odnosi samo na ciljani proizvod ili uslugu. Za ove potrebe se može primeniti klasifikacija prikupljenih tekstova kako bi se utvrdilo da li veb stranica sadrži diskusiju o relevantnom proizvodu/usluzi. Na osnovu odluke klasifikatora veb stranica se ili zadržava za dalju analizu ili se odbacuje. Neophodno je opskrbiti klasifikator listom relevantnih i irelevantnih pojmova na osnovu kojih će se vršiti procena sadržaja dokumenta. Klasifikator teksta se obučava primerima relevantnih i irelevantnih stranica ili komentara, a razvijeni model se kasnije može primeniti nad nepoznatim sajtovima kako bi utvrdio njihovu relevantnost. Dodatno, u ovom koraku se realizuje filtriranje i uklanjanje sadržaja koji nisu u tekstualnom formatu, poput HTML tagova, vrši se uklanjanje praznih polja, proširivanje skraćena, stemming (uklanjanje sufiksa) i uklanjanje stop reči.

*Ekstrakcija atributa* podrazumeva njihovo izdvajanje iz članka ili diskusije. Ukoliko su npr. predmet analize komentari o fakultetima, najbitniji atributi bi bili nazivi konkretnih fakulteta: Ekonomski fakultet, Pravni fakultet i drugo. Ukoliko nazivi (atributi) nisu adekvatno izdvojeni, kompletna analiza sentimenta ili mišljenja postaje irelevantna. Kao što je istaknuto u (Deshpande, et al., 2011) postoje tri osnovna pristupa ekstrakciji atributa:

1. Kreiranje rečnika ili taksonomija. Obezbeđuje listu relevantnih pojmova za analizirani domen. Na primer: za domen visokog obrazovanja potrebno bi bilo obezbediti taksonomiju svih poznatih fakulteta na osnovu koje bi se u preuzetim veb stranicama identifikovali relevantni sadržaji, koje bi sistem označio kao entitet *Fakultet*. Primena rečnika ne zahteva kompleksna podešavanja, međutim neophodna su česta ažuriranja, pogotovo kada je reč o sentiment analizi komentara o komercijalnim proizvodima kod kojih se često lansiraju novi modeli (mobilni telefoni, računari i slični proizvodi).
2. Formulisanje pravila. Tražene ključne reči obično ispoljavaju određeni šablon. U primeru fakulteta, pravilo bi se moglo napisati tako da omogućí označavanje svih alfanumeričkih karaktera koji se pojavljuju pre ili nakon stringa *fakultet*, npr. Fakultet organizacionih nauka i Medicinski fakultet. Ovaj pristup je robusniji od prethodnog, ali se pravila moraju manuelno ažurirati kako bi se obuhvatila sva odstupanja od predefinisano stringa pretrage.
3. Mašinsko učenje. Podrazumeva automatsko učenje pravila ekstrakovanja na osnovu datih primera za obučavanje u kojima su atributi na adekvatan način označeni. Pravila se uče formiranjem grafičkih i probabilističkih modela atributa i načina na koji su pojmovi oko njega uređeni.

*Ekstrakcija sentimenta* se najčešće realizuje upotrebom rečnika sentiment izraza<sup>25</sup> i njihove semantičke orijentacije. Ovaj pristup nosi određena ograničenja. Na primer, reč „visoka“ može

---

<sup>25</sup> U radu se *sentiment izraz* i *sentiment reč* koriste ravnopravno kako bi označili reči, fraze i idiome koje nose sentiment.

imati negativan sentiment kada se koristi u kontekstu školarine, dok u kontekstu ocene nosi pozitivan sentiment. Kada se u tekstu identifikuje relevantan entitet ili predefinisana sentiment reč, strukturirani sentiment se ekstrahuje iz rečenice u formi: entitet/reči i ocena. Pridružena ocena se odnosi na pozitivan ili negativan polaritet identifikovane reči u rečenici. Posebnu pažnju treba posvetiti negaciji. Tehnike obrade prirodnog jezika se koriste kako bi se utvrdili efekti negacije na pridruženu sentiment reč. Zavisno od strukture, kao i izvora prikupljenih podataka, u ovom koraku je potrebno izvršiti i procenu subjektivnosti iskazanog dokumenta, tako da se zadrže rečenice u kojima je prisutna subjektivnost, dok se one kojima se iskazuju činjenice odbacuju. Ovaj korak je potrebno realizovati u uslovima kada je izvor podataka forum, blog, društvene mreže, dok se na sajtovima za recenziranje uglavnom iskazuju subjektivni stavovi. Nakon ovog koraka, ekstrahovane sentiment reči su u strukturiranom formatu i mogu se učitati u bazu podataka radi daljih transformacija i izveštavanja.

*Klasifikacija sentimenta.* Zavisno od nivoa na kom se radi klasifikacija (dokument, rečenica, reč ili fraza, aspekt) iskazani sentiment se najčešće klasifikuje u pozitivnu ili negativnu kategoriju primenom različitih algoritama mašinskog učenja.

*Sumiranje sentimenta.* Kako bi ekstrahovani sentiment iz prethodnog koraka bio smislen i značajan za izveštavanje, potrebno ga je agregirati. Razlikuju se različiti nivoi agregiranja i sumiranja. U primeru fakulteta prirodan način agregiranja bilo bi grupisanje svih pozitivnih i negativnih reči prema pojedinačnim fakultetima, predmetima i drugim kriterijumima.

*Izveštavanje i vizualizacija rezultata* predstavljaju zadatke tradicionalne poslovne inteligencije i nisu predmet istraživanja ove doktorske disertacije.

## **3.2. Izazovi u sentiment analizi**

Brojni problemi i izazovi sentiment analize predstavljaju aktivne istraživačke oblasti. U ovom poglavlju je načinjen prikaz ključnih izazova i njihove relevantnosti u sentiment analizi.

### **3.2.1. Domenska i jezička zavisnost**

Potrošači iskazuju svoja mišljenja o temama iz različitih domena na forumima, blogovima, sajtovima za recenziranje, društvenim mrežama i mikroblogovima. Mišljenja i osećanja su iskazana na različite načine, različitim vokabularom, stilom pisanja, različito dužinom teksta i upotrebom žargona. Zavisnost od domena je delimično posledica promene u vokabularu, te identični izrazi mogu nositi potpuno drugačiji sentiment u različitim domenima. Na primer, nalog „Pročitaj knjigu“ može se tretirati kao pozitivan sentiment kada je reč o kritikama knjige, a negativan kada je reč o kritikama filma (Pang, et al., 2008). Klasifikator obučen skupom kritika o

jednom tipu proizvoda često ne pokazuje istu performantnost kada se primeni nad skupom kritika o drugom proizvodu ((Dave, et al., 2003), (Reinstein, et al., 2005)). Usled evidentne domenske zavisnosti, da bi se sentiment analiza mogla realizovati, potrebno je kreirati anotirane skupove podataka za svaki domen i razviti domen-specifičan klasifikator (Owsley, et al., 2006). Označeni podaci omogućavaju primenu različitih algoritama mašinskog učenja, pretežno nadgledanih klasifikacionih algoritama, kao i njihovu evaluaciju. Za veliki broj domena ne postoje razvijeni resursi, a njihovo kreiranje, pre svega anotiranje velike količine podataka, vremenski je zahtevno. Autori radova (Aue, et al., 2005) i (Blitzer, et al., 2007) bave se analizom mogućnosti prilagođavanja sistema za klasifikaciju sentimenta novim ciljnim domenima u uslovima nedostatka anotiranih korpusa. Prilagođavanje domena zahteva: a) prilagođavanje oznaka, jer jedna karakteristika koja se tretira kao pozitivna u izvornom domenu može imati suprotno značenje u ciljnom domenu, i b) prilagođavanje instance usled promene u vokabularu ili frekventnosti pojavljivanja reči u različitim domenima (Xia, et al., 2013).

S obzirom na zastupljenost engleskog jezika u društvenim medijima u vidu brojnih statusa i komentara, a pogotovo značaja koji ovaj jezik ima u naučnoj zajednici, razumljivo je da je i najveći broj resursa (rečnika sa sentiment oznakama i korpusa), kao i analitičkih pristupa i primenjenih tehnika i metoda analize sadržaja razvijen upravo za englesko govorno područje. Poslednjih godina prisutni su pokušaji za druge široko rasprostranjene jezike, poput kineskog, arapskog, nemačkog i španskog jezika ((Yunqing, et al., 2014), (Cambria, et al., 2012), (Remus, et al., 2010), (Waltinger, 2010), (Shalunts, et al., 2014), (Araque, et al., 2015), (Gamallo, et al., 2013), (Ibrahim, et al., 2015), (Refaee, et al., 2014), (Korayem, et al., 2012)). Usvajanje i prilagođavanje razvijenih resursa analizama sadržaja pisanih na drugom jeziku je usko povezano sa prilagođavanjem domenu (Pang, et al., 2008). Izrazi novog jezika se mogu uskladiti sa izrazima jezika za koji već postoje razvijeni resursi upotrebom višejezičkih rečnika (Mihalcea, et al., 2007), paralelnih korpusa (Kim, et al., 2006/2) ili mašinskog prevođenja (Bautin, et al., 2008).

### 3.2.2. Ekstrakcija atributa

U svojoj osnovi zadatak sentiment analize je klasifikacija iskazanog sentimenta. Svaki dokument ili rečenica, koji je predmet sentiment analize, najčešće se predstavlja kao vektor svih reči (atributa) koje se u njemu pojavljuju, te se prvi korak odnosi na ekstrakciju i odabir tekstualnih atributa. Jedna od osnovnih poteškoća se odnosi na visoku dimenzionalnost vektora obeležja kojima se obično opisuju tekstualni podaci. Svako jedinstveno obeležje koje se pojavljuje u kolekciji dokumenata predstavlja jednu dimenziju u vektorskom prostoru, a njihov broj može dostići i hiljade različitih obeležja (atributa). Visoka dimenzionalnost može dovesti do problema prekomerne obučenosti (engl. *overfitting*) prilagođavanjem klasifikatora specifičnom skupu za

obučavanje<sup>26</sup>. Iz tog razloga su metode odabira i ekstrakcije atributa važne kako bi se smanjila dimenzionalnost atributa, omogućila generalizacija i skratilo vreme za obučavanje (Varela, et al., 2013). Obeležja koja se učestalo koriste, opisana u radu (Medhat, et al., 2014), su:

1. *Prisutnost i frekvencija reči*. Atributi ovog tipa predstavljeni su jednom rečju ili n-gramima reči i brojem njihovog pojavljivanja u dokumentima. Pored merenja učestalosti pojavljivanja reči prebrojavanjem, može se posmatrati značaj reči koristeći meru *TF-IDF* koja meri težinu termina, odnosno stepen u kom reč doprinosi sadržaju dokumenta. *TF-IDF* se izračunava pomoću frekvencije termina u dokumentu *TF* (engl. *term frequency*) i inverzne frekvencije termina u skupu dokumenata *IDF* (engl. *inverse document frequency*), koja proverava broj dokumenata u kojima se pojavljuje konkretna reč (Weiss, et al., 2010). Najčešće upotrebljavane formule za izračunavanje *TF-IDF* mere su:

$$tf-idf(j) = tf(j) * idf(j)$$

$$idf(j) = \log\left(\frac{N}{df(j)}\right)$$

gde je  $j$  termin,  $tf(j)$  pokazuje broj pojavljivanja datog termina u dokumentu,  $df(j)$  je broj dokumenata u skupu dokumenata koji sadrže dati termin, a  $N$  je broj dokumenata u skupu.

Kada se reč pojavi u velikom broju dokumenata, smatra se nevažnom i pridružena težina je niža, blizu nuli. Kada je reč relativno jedinstvena i pojavljuje se u nekoliko dokumenata, pridružena težina se kreće bliže jedinici. Ovaj pristup je adekvatan u uslovima kada se ne zahteva dodatna lingvistička analiza teksta.

2. *Vrste reči* (engl. *POS – parts of speech*). Kada je cilj specifičniji, poput prepoznavanja imena, naziva mesta ili organizacije, potrebno je izvršiti dodatnu lingvističku analizu teksta i izvesti finije attribute. Tada se koristi automatsko određivanje vrste reči tzv. *POS tagging* (Weiss, et al., 2010). Pridevi predstavljaju važne indikatore osećanja, te je njihovo pronalaženje posebno bitan korak.
3. *Sentiment reči i fraze* predstavljaju reči koje se učestalo koriste u iskazivanju mišljenja, poput dobro, loše i sl. Pojedine fraze ukazuju na jasno mišljenje, stav ili osećanje, a da u njima nije upotrebljena konkretna sentiment reč (npr. *kapa dole*), što predstavlja poseban izazov.
4. *Negacija*. Pojava negacije u tekstu menja značenje i orijentaciju sentimenta, poput „*nije dobro*“, što je ekvivalentno lošem. Problem negacije je detaljnije opisan u narednom poglavlju.

<sup>26</sup> Ukoliko algoritam dostigne stepen *prekomerne obučenosti*, on će postizati izvisne rezultate za sve zadatke koji se mogu povući pod poznate šablone, ali neće davati dobre rezultate u zadacima koji se neznatno razlikuju.

### 3.2.3. Detekcija i rukovanje negacijom, sarkazmom i ironijom

Negacija predstavlja učestalu jezičku konstrukciju koja menja značenje reči, kao i orijentaciju iskazanog sentimenta i potrebno ju je uzeti u obzir pri analizi mišljenja i osećanja. Prisustvo negacije u rečenici ne znači da će se polaritet sentimenta uvek izmeniti. Naredni primeri ilustruju problematiku negacije.

(P.1.) *Dopada<sup>+</sup> mi se predmet.*

(P.2.) *[Ne dopada<sup>+</sup>] mi se predmet.*

(P.3.) *Ne samo da je predmet dosadan, nego su i materijali zastareli.*

(P.4.) *Ne dopada mi se predmet, ali su predavanja zanimljiva.*

U primeru P.1. iskazan je pozitivan stav sentiment izrazom *dopada*. Negacija ove rečenice, P.2, upotrebom ključne reči *ne* menja njeno značenje i iskazuje negativan stav. U primeru P.3. negacija ne menja sveukupni sentiment rečenice koja ostaje negativna. Nadalje, prisustvo negacije u rečenici ne znači da se svako iskazano mišljenje negira, odnosno da se njihovo značenje menja. U primeru P.4. negacija menja polaritet sentimenta u prvom delu rečenice, dok se drugim delom rečenice iskazuje pozitivan stav.

Kada se rukuje negacijom, potrebno je ispravno identifikovati njen opseg važenja, odnosno koji deo rečenice će imati izmenjeno značenje usled prisustva negacije (Wiegand, et al., 2010). U većini slučajeva upotreba negacije nije jednostavna, kao u primeru P.2. Pored standardnih ključnih reči negacije (*ne*, *nije*, *nema* i sl.) vrlo često se koriste neutralizatori (reči koje umanjuju značenje poput *malo*, *donekle* i sl.), veznici i glagoli u potencijalu koji negiraju značenje ili umanjuju iskazani sentiment.

Negacija se može iskazati i na vrlo suptilan način upotrebom sarkazma i ironije, koji se prilično teško identifikuju. Neki primeri njihove upotrebe su:

(P.5) *Jedna žena koja je jako posvećena svom poslu. Toliko posvećena da održava konsultacije jednom mesečno a i kad ih održi nigde je nema :).*

(P.6) *Zanimljiva osoba... Ako niste žensko, onda vam da 3 ocene manje u startu.*

Sarkazam i ironija su uobičajeni u recenzijama proizvoda i usluga, a učestalo se pojavljuju i u političkim diskusijama i komentarima. Kako se navodi u radu (Filatova, 2012), jedan od osnovnih problema u vezi sa zadatkom identifikovanja ironije je nepostojanje saglasnosti među istraživačima po pitanju formalne definicije ironije ili sarkazma i njihove strukture. Smatra se da je primena ironije i sarkazma, kao i termini koji ukazuju na njih, promenjiva i da sarkazam ima

regionalne varijacije. Iz tog razloga nije moguće definisati jasna uputstva za identifikovanje ironije i sarkazma. Međutim, ljudi generalno razumeju značenje ironičnih izraza i mogu ih pouzdano identifikovati. Dodatno, autorka navodi da se određena pojavljivanja sarkastičnih izjava mogu shvatiti samo kada su smeštena u određeni kontekst. Zadaci identifikovanja negacije, sarkazma i ironije predstavljaju bitan korak sentiment analize, jer njihovo ispravno detektovanje poboljšava performanse sistema analize sentimenta.

### **3.2.4. Utvrđivanje polariteta**

Upotreba rečnika polarizovanih (pozitivnih i negativnih) reči i fraza predstavlja uobičajeni pristup sentiment analizi (Wilson, et al., 2005/2), na čemu je bazirano i istraživanje koje je predmet doktorske disertacije. Ovakvi rečnici se mogu kreirati manuelno ili automatskim ekstrahovanjem pojmova. Generisanje polariteta reči i fraza predstavlja aktivni istraživački pravac, a predložene su razne tehnike za učenje polariteta, uključujući tehnike bazirane na korpusu, koje podrazumevaju manuelno označavanje pozitivnih i negativnih reči i tehnike koje koriste informacije o leksičkim odnosima i resurse poput *WordNet* mreže pojmova na osnovu kojih se kreira sistem za automatsko ekstrahovanje pojmova (Wilson, et al., 2005/2). Pojedine reči posmatrane samostalno mogu nositi jedan polaritet koji će se izmeniti kada su one upotrebljene u određenom kontekstu. Ovaj problem predstavlja poseban izazov u klasifikaciji sentimenta na nivou fraze. Pored negacije, na promenu kontekstualnog polariteta utiče i samo značenje reči i njihova upotreba, sintaksička uloga reči u rečenici, kao i reči koje umanjuju sentiment (npr. reč malo) (Wilson, 2008). Autori radova ((Wilson, et al., 2005/2), (Wilson, 2008), (Agarwal, et al., 2009)) rešavaju problem identifikovanja kontekstualnog polariteta anotiranjem (označavanjem) korpusa.

### **3.2.5. Utvrđivanje subjektivnosti dokumenta**

Nasuprot objektivnim iskazima koji odražavaju činjenice, subjektivni iskazuju lična osećanja, mišljenja i uverenja. Subjektivni iskazi se mogu pojaviti u formi mišljenja, želje, verovanja, sumnje, spekulacije, osude, navoda (Liu, 2012). Mnogi od njih podrazumevaju sentiment, ali u određenim situacijama subjektivni iskazi ne sadrže sentiment (npr. Želim foto-aparat koji će praviti kvalitetne fotografije). Identifikovanje subjektivnosti svodi se na problem klasifikacije tekstova na one koji sadrže činjenice i one koji sadrže subjektivne iskaze. Nad izdvojenim subjektivnim iskazima se može sprovesti sentiment analiza kako bi se izvršila klasifikacija sadržaja prema polaritetu. Iz tog razloga, u radu (Wiegand, et al., 2009) ističu da je potrebno napraviti distinkciju između detektovanja subjektivnosti i sentiment analize, jer predstavljaju dva različita zadatka obrade prirodnog teksta. Sentiment analiza se može realizovati samostalno ili u kombinaciji sa detektovanjem subjektivnosti. Prepoznavanje subjektivnosti i uklanjanje

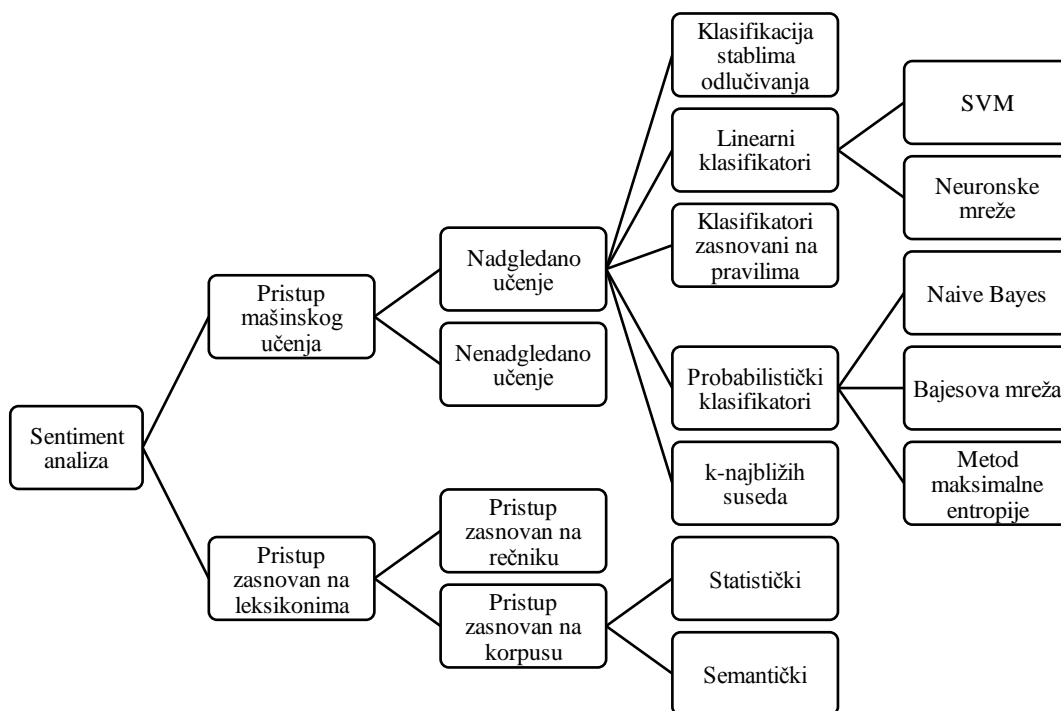
objektivnih rečenica iz tekstova pre sprovođenja sentiment analize daje višu tačnost klasifikacije (Pang, et al., 2004).

### 3.3. Metode, tehnike i pristupi u sentiment analizi

Zadatak sentiment analize u svojoj osnovi predstavlja zadatak klasifikacije iskazanih osećanja, najčešće na pozitivne i negativne. Tehnike klasifikacije sentimenta mogu se u osnovi podeliti na *tehnike mašinskog učenja*, koje koriste lingvističke attribute, *pristup zasnovan na leksikonima* (rečnici ili korpus), koji kao input prihvata listu poznatih i prethodno pripremljenih sentiment izraza, i *hibridni pristup*, koji kombinuje prethodna dva. Slika 3.1. daje prikaz najčešće korišćenih tehnika klasifikacije sentimenta detaljnije opisanih u narednim poglavljima.

Tehnike mašinskog učenja se dele na nadgledano i nenadgledano učenje. Metodi nadgledanog učenja zahtevaju veliku količinu označenih podataka za obučavanje, dok se metodi nenadgledanog učenja koriste u uslovima nepostojanja pripremljenih podataka.

Pristupi zasnovani na leksikonima su bazirani na sentiment rečnicima na osnovu kojih prepoznaju sentiment u tekstovima. Pristup zasnovan na rečnicima polazi od osnovnih sentiment reči i potom u rečniku pronalazi sinonime i antonime. Pristupi zasnovani na korpusu polaze od liste sentiment reči na osnovu kojih se pronalaze druge sentiment reči u velikom korpusu uz njihovu kontekstnu orijentaciju (Medhat, et al., 2014).



Slika 3.1. Tehnike klasifikacije sentimenta (Medhat, et al., 2014)

### 3.3.1. Algoritmi nadgledanog učenja u sentiment analizi

Klasifikatori nadgledanog učenja se automatski razvijaju učeći karakteristike kategorija na osnovu kolekcije prethodno klasifikovanih i označenih dokumenata za obučavanje. U izgradnji klasifikacionih modela nadgledanim učenjem potrebno je razmotriti četiri osnovna pitanja (Feldman, et al., 2013):

- Doneti odluku o kategorijama koje će se koristiti u klasifikaciji instanci.
- Obezbediti skupove za obučavanje za svaku kategoriju.
- Doneti odluku o atributima koji opisuju svaku instancu. Kako većina algoritama nadgledanog učenja ima sposobnost fokusiranja na relevantne attribute, uputno je generisati što više atributa.
- Doneti odluku o algoritmu koji će se primeniti.

U nastavku poglavlja opisane su osnove najčešće korišćenih algoritama nadgledanog učenja.

#### 3.3.1.1. Stabla odlučivanja

Većina klasifikatora rezultuje vrednostima koje ljudi bez poznavanja pozadinskog mehanizma rada algoritma teško interpretiraju i shvataju. Simbolički klasifikatori, od kojih su stabla odlučivanja najpoznatija tehnika, prevazilaze ovaj problem. Ona na hijerarhijski način dekomponuju podatke za obučavanje predstavljajući ih čvorovima stabla koji su stavljeni u fizički odnos pomoću potega, odnosno uslova koje atributi moraju zadovoljiti. Svi krajnji čvorovi, takozvani listovi stabla, predstavljaju konačni cilj ili klasu kojoj dokument pripada.

Proces klasifikacije stablima odlučivanja polazi od korena stabla. Vršni se pretraga stabla sleva nadesno, krećući se iz svakog čvora na naredni nivo hijerarhije, do novog čvora, onom granom koja zadovoljava definisani uslov (Bošnjak, 2006). Za primenu stabla odlučivanja potrebno je obezbediti veliku količinu primera za obučavanje koji su opisani diskretizovanim atributima, kao i predefinisane klase kojima primeri pripadaju (Xu, et al., 2011). Postoje tri načina particionisanja stabla odlučivanja. Prvi način ispituje prisustvo ili odsustvo jedne ili više reči u dokumentu, drugi je baziran na frekventnosti reči i sličnosti dokumenata, dok treći pristup koristi diskriminator za particionisanje stabla, poput Fišerove linearne diskriminacione funkcije (Chakrabarti, et al., 2003). Najčešće korišćeni algoritmi za klasifikaciju teksta stablima odlučivanja predstavljaju ID3 i C4.5, kao i njihove varijacije ((Medhat, et al., 2014), (Xu, et al., 2011)).



### 3.3.1.2. Linearni klasifikatori: Neuronske mreže i Mašine potpornog vektora

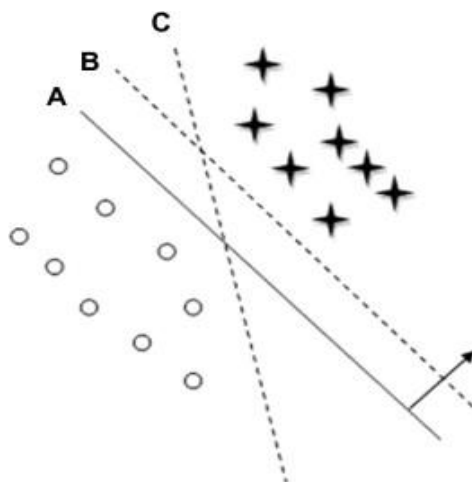
Neuronske mreže se sastoje iz skupa međusobno povezanih neurona koji primaju inpute i čine ulazni sloj mreže. Neuroni koji primaju signale samo od drugih neurona sačinjavaju jedan ili više skrivenih slojeva mreže. Svaki sloj mreže prima ili impulse u vidu vektora podataka ili izlaze iz prethodnih slojeva mreže i paralelno ih obrađuje. Skup izlaznih neurona koji daje finalni rezultat rada mreže naziva se izlaznim slojem. Znanje koje mreža nauči pohranjuje se u formi težinskih faktora pridruženih konekcijama među čvorovima mreže (Bošnjak, et al., 2010). U klasifikaciji dokumenata koriste se *back-propagation* neuronske mreže sa ERBP (engl. *error back propagation*) algoritmom, gde dokumenta za obučavanje prihvataju input čvorovi, a u slučaju pogrešne klasifikacije greška se prostire unazad kroz mrežu modifikujući težinske faktore sa ciljem minimiziranja greške (Feldman, et al., 2013). Najjednostavniji vid neuronskih mreža je sa dva sloja, ulaznim i izlaznim. Ovakve mreže su ekvivalentne linearnom klasifikatoru. Formalni zapis linearne funkcije neuronske mreže je (Medhat, et al., 2014):

$$p_i = A \cdot \bar{X}_i$$

gde  $A$  predstavlja skup težinskih faktora,  $\bar{X}_i$  ulazne neurone (frekventnost reči u  $i$ -tom dokumentu), a  $p_i$  je rezultujuća klasa.

Iako postoje kompleksnije mreže, sa više skrivenih slojeva, praksa je pokazala da se u klasifikovanju dokumenata najbolje performanse postižu upravo najjednostavnijom implementacijom ((Van de Camp, et al., 2012), (Feldman, et al., 2013)). Neuronske mreže daju dobre rezultate u kompleksnim, nelinearnim domenima, što prema autorima rada (Xu, et al., 2011) predstavlja njihovu prednost, dok kao nedostatak navode dug proces obučavanja i činjenicu da je modelovano znanje implicitno, skriveno u strukturi mreže te se teško interpretira.

Metoda *potpornih vektora* – SVM (engl. *Support vector machines*) predstavlja efikasan metod koji problem klasifikacije svodi na pronalaženje hiperravni separacije. Hiperravan maksimalno razdvaja podatke u prostoru na pozitivne i negativne instance na osnovu klasnog atributa ((Feldman, et al., 2013), (Tan, et al., 2006)). Na slici 3.2. definisane su dve klase,  $x$  i  $o$ , i tri hiperravni, A, B, C. Hiperravan A daje najbolju separaciju klasa jer je udaljenost tačaka najviša, što ujedno čini i maksimalnu marginu separacije. Odluka o granicama separacije definisana je preko podskupa primera iz skupa za obučavanje, tzv. potpornih vektora (engl. *support vectors*) ((Tan, et al., 2006), (Feldman, et al., 2011)). Prednost SVM algoritma se ogleda u njegovoj mogućnosti klasifikovanja podataka velike dimenzionalnosti.



Slika 3.2. Upotreba SVM algoritma u klasifikaciji teksta (Medhat, et al., 2014)

### 3.3.1.3. Klasifikator zasnovan na pravilima

Metode zasnovane na pravilima modeluju podatke u prostoru pomoću seta pravila. Pravila su u disjunktivnoj normalnoj formi (Berlanga, et al., 2006):

$$\text{If } X_1 = \hat{A}_1 \text{ AND } \dots \text{ AND } X_n = \hat{A}_n \text{ Then } Y = C \text{ sa CF}$$

gde svaka input varijabla  $X_i$  kao vrednost prihvata set termina ili lingvističkih oznaka  $\hat{A}_i = \{A_{i1} \text{ OR } \dots \text{ OR } A_{iLi}\}$  povezanih disjunktivnim operatorom, dok output varijabla (Y) poprima jednu od vrednosti klasa. Uslovi koji se testiraju pravilima odnose se na prisustvo određenog termina u dokumentima za obučavanje. Pravilo takođe uključuje i faktor pouzdanosti (CF  $\in [0, 1]$ ) koji predstavlja pouzdanost klasifikacije u klasu predstavljenu konsekvantom pravila.

Iz seta pravila koji tačno klasifikuju sve primere za obučavanje, metod učenja selektuje najbolja pravila (Feldman, et al., 2013).

### 3.3.1.4. Probabilistički klasifikatori

Najzastupljeniji probabilistički klasifikatori teksta su Naïve Bayes, Bayesove mreže i klasifikator maksimalne entropije (engl. *Maximum Entropy*). U kontekstu klasifikacije teksta, oni su opisani u radu (Medhat, et al., 2014) na sledeći način:

Bayesov klasifikator predstavlja jednostavan i najčešće korišćeni metod za klasifikaciju teksta. Bazira se na teoriji verovatnoće, odnosno na Bayesovoj teoremi. *Naïve Bayes klasifikator* određuje aposteriornu verovatnoću klase na osnovu distribucije reči u dokumentu. Koristi se

reprezentacija dokumenta korpom reči (engl. *Bag of Words*) koja rezultuje listom sentiment reči, a ignoriše njihovu poziciju u dokumentu. Koristeći Bayesovu teoremu predviđa se verovatnoća da će određeni set atributa pripadati nekoj klasi.

$$P(klasa|atributi) = \frac{P(klasa) * P(atributi|klasa)}{P(atributi)}$$

$P(klasa)$  je apriorna verovatnoća ili marginalna verovatnoća klase.  $P(atributi|klasa)$  je apriorna verovatnoća klasifikacije seta atributa u datu klasu.  $P(atributi)$  predstavlja apriornu verovatnoću da će se dati set atributa pojaviti i igra ulogu konstante normalizacije. Kako je polazna pretpostavka *Naive Bayes* klasifikatora nezavisnost atributa, jednačina se može preformulisati na sledeći način:

$$P(klasa|atributi) = \frac{P(klasa) * P(f_1|klasa) * ... * P(f_n|klasa)}{P(atributi)}$$

Ukoliko se pretpostavi da su svi atributi potpuno zavisni, dobija se *Bayesova mreža* koja predstavlja usmereni aciklični graf čiji su čvorovi slučajne promenljive, a ivice pokazuju zavisnost promenljivih. Kako Bayesove mreže predstavljaju kompletan model varijabli i njihovih veza, modelom se definiše sveukupna uslovna verovatnoća za sve varijable (čvorove). Kompleksnost primene i razlog zbog kog se ne koriste često u praksi za klasifikaciju teksta ogleda se u računarskoj kompleksnosti i činjenici da je i dalje potrebno definisati veliku količinu brojki ((Burde, et al., 2015), (Medhat, et al., 2014), (Wu, et al., 2005)).

*Klasifikator maksimalne entropije* (ME), za razliku od *Naive Bayes*, ne polazi od pretpostavke da su atributi nezavisni, nego koristi optimizaciju zasnovanu na pretrazi kako bi pronašao težine svakog atributa, koji se potom mogu kombinovati kako bi se utvrdila najverovatnija klasa za dati set atributa. Verovatnoća svake klase se izračunava na osnovu sledeće formule (Pang, et al., 2002):

$$P_{ME}(c | d) := \frac{1}{Z(d)} \exp\left(\sum_i \lambda_{i,c} F_{i,c}(d, c)\right)$$

gde je  $Z(d)$  funkcija normalizacije, a  $F_{i,c}$  je funkcija atributa/klase za atribut  $f_i$  i klasu  $c$ , definisana na sledeći način:

$$F_{i,c}(d, c') := \begin{cases} 1, & n_i(d) > 0 \text{ i } c' = c \\ 0, & \text{u suprotnom} \end{cases}$$

Pri tome je  $\{f_1, \dots, f_m\}$  predefinisani set  $m$  atributa koji se može pojaviti u dokumentu (npr. reči poput *lepo*, *nije dobro* i *sl.*),  $n_i(d)$  označava broj pojavljivanja reči  $f_i$  u dokumentu  $d$ . Svaki dokument  $d$  je predstavljen vektorom dokumenta  $\sim d := (n_1(d), n_2(d), \dots, n_m(d))$ . Tako se određena funkcija atributa/klase može aktivirati ako i samo ako se npr. bigram „*nije dobro*“ pojavi uz pretpostavku da je sentiment dokumenta negativan. Težine atributa su u oznaci  $\lambda_{i,c}$ . Prema definiciji  $P_{ME}$  više vrednosti  $\lambda_{i,c}$  govore da je  $f_i$  jači indikator klase  $c$ .

Naïve Bayes algoritam i klasifikator maksimalne entropije su široko primenjeni u klasifikaciji tekstova, a klasifikator maksimalne entropije ponekad čak nadmašuje performanse Naïve Bayesa (Pang, et al., 2002).

### 3.3.1.5. Klasifikator $k$ -najbližih suseda

Klasifikacija teksta zasnovana na algoritmu  $k$ -najbližih suseda (engl. *k-Nearest Neighbor*, u daljem tekstu *kNN*) pokazala se performantnom i doslednom po pitanju kvaliteta rezultata (Xu, et al., 2011). U ovom pristupu se ne gradi klasifikacioni model, nego se kategorizacija dokumenata vrši na osnovu sličnosti novog primera i primera u skupu za obučavanje. Radi ocene sličnosti uzoraka, mogu se koristiti sve dostupne mere poput Euklidskog rastojanja, kosinusne sličnosti (engl. *cosine similarity*) i sl. Metrika se odabira na osnovu tipa podataka koji se klasifikuju.

### 3.3.2. Primena nenadgledanog učenja u sentiment analizi

Za realizaciju klasifikacije teksta potrebno je pripremiti veliku količinu označenih dokumenata koji će se upotrebiti u obučavanju algoritama (nadgledane) klasifikacije. U realnim uslovima je jednostavno prikupiti veliku količinu podataka sa interneta, dok je teško pripremiti dovoljan broj jasno razgraničenih i označenih dokumenata (npr. dokumenti koji sadrže pozitivne ili negativne komentare). Sentiment reči se mogu upotrebiti u klasifikaciji sentimenta nenadgledanim učenjem jer predstavljaju ključan faktor u klasifikaciji dokumenata.

Autor rada (Turney, 2002) predlaže klasifikaciju na osnovu fiksnih sintaksnih paterna koji su identifikovani u iskazivanju sentimenta. Sintakсни paterni su definisani na osnovu POS tagova i odgovaraju sadržajima koji su pisani na engleskom jeziku. Algoritam je sačinjen iz tri koraka. U prvom koraku se ekstrahuju parovi uzastopnih reči koji odgovaraju predefinisanim paternima. Predefinisani paterni za engleski jezik su:

$\text{Pride}_{V(\text{prva reč})} + \text{Imenica/Množina imenice}_{(dru\text{ga reč})} + \text{Bilo koja vrsta reči}_{(\text{treća reč})}$

$\text{Prilo\text{g}}_{(\text{prva reč})} + \text{Pride}_{V(\text{dru\text{ga reč})} + \text{Imenica}_{(\text{treća reč})}$

Pridev<sub>(prva reč)</sub> + Pridev<sub>(druga reč)</sub> + Imenica<sub>(treća reč)</sub>

Imenica<sub>(prva reč)</sub> + Pridev<sub>(druga reč)</sub> + Imenica<sub>(treća reč)</sub>

Prilog<sub>(prva reč)</sub> + Glagol<sub>(druga reč)</sub> + Bilo koja vrsta reči<sub>(treća reč)</sub>

U primeru drugog paterna, dve konsekventne reči će se ekstrahovati ukoliko nakon priloga i prideva sledi imenica. Treća reč se ne ekstrahuje, ali se uzima kao kriterijum. Pridevi, prilozi i njihovi komparativi i superlativi su vrste reči kojima je najčešće iskazano mišljenje. Međutim, oni mogu nositi različit sentiment, zavisno od konteksta u kom su iskazani, a imenice i glagoli ukazuju na kontekst (Liu, 2012). Primer: pridev *nepredvidiv* može imati negativnu konotaciju u recenziji automobila „*menjač je nepredvidiv*“, dok može nositi pozitivan sentiment u recenziji filмова „*nepredvidiva radnja*“.

U sledećem koraku se procenjuje sentiment orijentacija (SO) ekstrahovanih fraza PMI (engl. *point-wise mutual information*) merom:

$$PMI(pojam_1, pojam_2) = \log_2 \left( \frac{\Pr(pojam_1 \wedge pojam_2)}{\Pr(pojam_1) \Pr(pojam_2)} \right)$$

PMI meri stepen statističke zavisnosti između dva pojma, gde  $\Pr(pojam_1 \wedge pojam_2)$  predstavlja stvarnu verovatnoću zajedničkog pojavljivanja pojma<sub>1</sub> i pojma<sub>2</sub>, dok  $\Pr(pojam_1) \Pr(pojam_2)$  predstavlja verovatnoću zajedničkog pojavljivanja dva pojma ukoliko su statistički nezavisni. Orijentacija sentimenta fraze se kalkuliše na osnovu njene povezanosti sa referentnom pozitivnom reči „*odličan*“, odnosno referentnom negativnom reči „*loš*“:

$$SO(fraza) = PMI(fraza, "odličan") - PMI(fraza, "loš")$$

Povezanost referentne reči i fraze proverava se davanjem upita pretraživaču veba<sup>27</sup> i prikupljanjem broja pogodaka (*hits*). Tako se verovatnoće u prvoj jednačini mogu proceniti pretraživanjem pojmova zajedno i odvojeno. U istraživanju prezentovanom u radu (Turney, 2002) koristi se NEAR operator pretraživača. Prethodna formula je modifikovana uvođenjem broja pogodaka:

$$SO(fraza) = \log_2 \left( \frac{hits(fraza \text{ NEAR } "odličan") \cdot hits("loš")}{hits(fraza \text{ NEAR } "loš") \cdot hits("odličan")} \right)$$

<sup>27</sup> Autor rada (Turney, 2002) koristi AltaVista napredni pretraživač veba. Za razliku od Google veb pretraživača, AltaVista nudi operator pretraživanja *Near*.

Algoritam kalkuliše prosečnu *SO* svih fraza u dokumentu i klasifikuje dokument kao pozitivan ukoliko je *SO* pozitivno, u suprotnom – kao negativan.

Pored opisanih pristupa još jedan uobičajeni nenadgledani pristup je klasifikacija zasnovana na leksikonima. Detaljniji opis dat je u narednom poglavlju.

### 3.3.3. Klasifikacija zasnovana na leksikonima

Sentiment reči se koriste u mnogim zadacima klasifikacije. Pozitivnim rečima i frazama iskazuju se željena stanja, dok se negativnim iskazuju neželjena, i njihova objedinjena lista naziva se leksikonom mišljenja ili sentimenta. Postoje tri osnovna načina njihovog prikupljanja i sačinjavanja. *Manuelni pristup* zahteva intenzivan rad i vremenski je zahtevan. Retko se primenjuje samostalno, već kao kombinacija sa automatizovanim pristupom u svrhu završne provere kako bi se izbegle greške kojima rezultuju automatizovani pristupi (Medhat, et al., 2014).

*Pristup zasnovan na rečnicima* podrazumeva upotrebu rečnika za kompajliranje liste sentiment reči. Većina rečnika daje sinonime i antonime za svaku reč. Kreiranje rečnika polazi od nekoliko osnovnih reči i primenjuje se *bootstrapping* tehnika na osnovu strukture sinonima i antonima u rečniku. Polazni set reči se prikuplja manuelno i pridružuje im se sentiment orijentacija. Algoritam zatim uvećava set reči pretražujući sinonime i antonime u *WordNet*-u ili drugom onlajn rečniku. Nove reči se dodaju polaznoj listi reči i započinje se nova iteracija. Iterativni proces se okončava kada nema novih reči. Rezultujuća lista reči može imati brojne greške, te je potrebno izvršiti manuelnu korekciju liste kako bi se korigovale i otklonile greške. Iako je vremenski zahtevna (nekoliko dana za osobu kojoj je maternji jezik jezik na kojem su pisane sentiment reči (Liu, 2012)), primena manuelne provere efikasno filtrira rečnik. Prednost primene pristupa zasnovanog na rečniku je pronalaženje velikog broja sentiment reči sa njihovim orijentacijama. Osnovni nedostatak ovog pristupa se ogleda u njegovoj nemogućnosti pronalaženja sentiment reči i njihove orijentacije koje su specifične kontekstu ili uzanom domenu.

*Pristup zasnovan na korpusu* omogućuje identifikovanje sentiment reči specifičnih za domen koji je predmet analize ili za određeni kontekst. Međutim, pristup nije efikasan kao prethodni zbog nemogućnosti obuhvatanja svih reči u jeziku, a domen-specifičan sentiment rečnik je često nedovoljan, kako u istom domenu pojedine reči zavisno od konteksta mogu poprimiti i pozitivnu i negativnu orijentaciju sentimenta. Iz tog razloga autori u radu (Ding, et al., 2008) predlažu pristup kojim se ove reči povezuju sa aspektom/temom rečenice, što se pokazalo efikasnijim pristupom. Pristup zasnovan na korpusu se primenjuje u dva osnovna scenarija: 1. na osnovu polazne liste opštih sentiment reči otkrivaju se ostale sentiment reči i njihova orijentacija u domenskom korpusu i 2. prilagođavanje opšteg sentiment leksikona novom korišćenjem domenskog korpusa za primenu sentiment analize u datom domenu (Liu, 2012). Baziran je na

sintaksnim paternima ili paternima koji se pojavljuju zajedno sa sentiment rečima iz bazične liste kako bi se pronašle ostale sentiment reči u korpusu. Paterni se mogu pronaći upotrebom statističkih tehnika. Autori u radu (Medhat, et al., 2014) navode neke od pristupa: naknadno deriviranje polariteta na osnovu pojave prideva u nizu (npr. pridevi povezani veznikom i imaju isti polaritet, dok pridevi povezani veznicima ili, ali ili rečcom međutim, obično imaju suprotan sentiment polaritet), analiza učestalosti pojavljivanja reči u anotiranom korpusu (ukoliko se reč češće pojavljuje u pozitivnim tekstovima, njen polaritet je pozitivan, i obrnuto), analiza povezanih reči koje se učestalo pojavljuju u korpusu (ukazuje da one najverovatnije imaju isti polaritet, a polaritet nepoznatih reči se određuje računanjem relativne frekvencije njihovog pojavljivanja sa drugim rečima). Semantički pristup podrazumeva direktno dodeljivanje sentiment vrednosti rečima. Semantički bliskim rečima dodeljuje se ista sentiment vrednost.

### 3.3.4. Sentiment analiza zasnovana na aspektima

Klasifikacija teksta na nivou dokumenta i rečenice daje broj potrošača koji su zadovoljni ili nezadovoljni proizvodom, na osnovu čega se mogu identifikovati trendovi u percepciji potrošača, međutim, ovakva klasifikacija je često nedovoljna jer ne daje tačne razloge (ne)zadovoljstva. Takođe, standardnim pristupom klasifikaciji ne identifikuju se ciljevi ka kojima su mišljenja i iskazani sentimenti usmereni, niti se pridružuje sentiment tim ciljevima. Ako se pretpostavi da svaki dokument evaluira jedan entitet, pozitivni dokumenti ne govore da autor ima pozitivan stav o svim aspektima datog entiteta, i obrnuto. Aspektno orijentisana analiza sentimenta pruža sveobuhvatniju analizu sentimenta prema različitim aspektima predmeta analize.

Za realizaciju aspektno orijentisane sentiment analize potrebno je ekstrahovati sve relevantne aspekte, a potom izvršiti klasifikaciju osećanja prema aspektima kako je opisano u radu (Liu, 2012):

**Ekstrakcija evaluiranih aspekata** se može tretirati i kao zadatak ekstrakcije informacija. Mišljenje je uvek usmereno ka nekom cilju koji najčešće predstavlja aspekt ili temu koju treba preuzeti iz rečenice. Postoje četiri osnovna pristupa ekstrakciji aspekata:

1. *Ekstrakcija zasnovana na učestalim imenicama i imenskim frazama*, koja zahteva da se nad dokumentima primeni POS označavanje kako bi se na osnovu oznaka identifikovale imenice u tekstovima. Vršiti se prebrojavanje njihovog pojavljivanja. Imenice ili fraze koje se učestalo pominju predstavljaju bitne aspekte entiteta o kojem se diskutuje. Reč je o jednostavnom, ali efikasnom pristupu, koji se široko primenjuje u praksi komercijalnih kompanija.
2. *Ekstrakcija na osnovu odnosa mišljenja i cilja* vrši se na osnovu poznatih sentiment reči. Ista sentiment reč se može upotrebiti da opiše različite aspekte. Ukoliko rečenica ne

sadrži aspekt koji se učestalo pojavljuje u domenu, a sadrži neku sentiment reč, ekstrahovaće se najbliža imenica ili imenska fraza (Hu, et al., 2004). Primer ovakve rečenice: *Softver je odličan*. Reč *odličan* je poznata sentiment reč. Na osnovu njenog prisustva u rečenici reč *softver* se izdvaja kao aspekt.

3. Pored *algoritama nadgledanog učenja* opisanih u [Poglavlju 3.3.1](#), u ekstrahovanju aspekata se učestalo koriste metodi sekvencijalnog učenja Skriveni Markovljevi modeli (engl. *Hidden Markov models* – HMM) i CRF (engl. *Conditional Random Fields*). Oba pristupa zahtevaju manuelno označene podatke za obučavanje.
4. *Ekstrakcija modelovanjem tema* predstavlja metod nenadgledanog učenja zasnovan na verovatnoći. Modelovanje tema rezultuje nizom reči. Svaki klaster formira jednu temu i predstavlja distribuciju verovatnoće pojave reči u kolekciji dokumenata. Postoje dva osnovna modela kojima se modeluju teme: pLSA (engl. *Probabilistic Latent Semantic Analysis*) i LDA (engl. *Latent Dirichlet allocation*). Modeli tema se mogu proširiti i na modelovanje brojnih drugih tipova informacija. U kontekstu sentiment analize, modelom se u isto vreme mogu obuhvatiti sentiment reči i teme/aspekti. U radu (Liu, 2012) autor ističe da je primena modelovanja tema u aplikacijama sentiment analize značajno limitirana njihovom potrebom za velikom količinom podataka i značajnim podešavanjima potrebnim radi postizanja prihvatljivih rezultata. Ovim pristupom se sa lakoćom mogu prepoznati uopštene i učestale teme u velikim kolekcijama podataka, ali se ne mogu sa lakoćom identifikovati lokalno frekventni, ali globalno retki aspekti, koji su najkorisniji u aplikacijama jer su najrelevantniji za specifičan entitet za koji je korisnik zainteresovan (Liu, 2012).

**Klasifikacija sentimenta prema aspektima** određuje polaritet iskazanih osećanja prema svakom aspektu. Može se realizovati metodama nadgledanog učenja koje se koriste u klasifikaciji na nivou rečenice ili reči, kao i na osnovu rečnika. Nadgledano učenje zavisi od podataka za obučavanje. Klasifikator obučen nad podacima iz jednog domena često ne daje dobre rezultate u drugom domenu. S druge strane, pristup baziran na rečnicima daje dobre rezultate u velikom broju domena (Liu, 2012). Ovaj pristup se zasniva na upotrebi sentiment rečenika koji sadrže listu sentiment reči, fraza i idioma, kompozitne izraze i pravila mišljenja sa ciljem utvrđivanja orijentacije sentimenta prema svakom aspektu u analiziranoj rečenici. Uzima u obzir i reči koje menjaju sentiment (engl. *shifters*), rečenice koje sadrže veznik *ali*, kao i druge konstrukte koji utiču na polaritet sentimenta. U uslovima kada su poznati aspekti, autori rada (Ding, et al., 2008) predlažu primenu metode koja podrazumeva: **1) Označavanje sentiment reči i fraza** u svakoj rečenici koja sadrži jedan ili više aspekata. Svakoj pozitivnoj reči se pridružuje ocena sentimenta +1, dok svakoj negativnoj reči ocena -1. **2) Primena promene sentimenta** usled pojave reči ili fraza koje menjaju polaritet sentimenta, poput negacije. **3) Rukovanje suprotnim naporednim odnosom**. Prisustvo veznika *ali* ili nekih reči ukazuje na suprotnost nezavisnih predikatskih rečenica i na promenu sentiment orijentacije. Na rečenice koje sadrže suprotni veznik (najčešće *ali*) primenjuje se sledeće pravilo: sentiment orijentacija pre i nakon pojave suprotnog veznika je suprotna ukoliko se mišljenje na jednoj strani ne može utvrditi. Uslov u pravilu je postavljen



zbog činjenice da suprotni veznici ne menjaju uvek sentiment, npr. „*Model laptopa X je dobar, ali je model laptopa Y bolji.*“. 4) *Agregiranje mišljenja* podrazumeva primenu funkcije agregiranja na rezultujuće ocene sentimenta kako bi se utvrdila konačna orijentacija sentimenta za svaki aspekt u rečenici. Ukoliko rečenica  $s$  sadrži set aspekata  $\{a_1, \dots, a_m\}$  i set sentiment reči  $\{sw_1, \dots, sw_n\}$ , sa njihovim ocenama dobijenim u prethodnim koracima, sentiment orijentacija svakog aspekta  $a_i$  u  $s$  je definisana sledećom agregacionom funkcijom:

$$score(a_i, s) = \sum_{sw_j \in s} \frac{sw_j\_so}{dist(sw_j, a_i)}$$

gde je  $sw_j$  sentiment reč/fraza u  $s$ ,  $dist(sw_j, a_i)$  je udaljenost između aspekta  $a_i$  i sentiment reči  $sw_j$  u  $s$ , a  $sw_j\_so$  je ocena sentimenta za  $sw_j$ . Sentiment rečima koje su više udaljene od aspekta  $a_i$  dodeljuju se niže težine. Ukoliko je konačna ocena pozitivna, onda je mišljenje o aspektu  $a_i$  u  $s$  pozitivno. Ukoliko je konačna ocena negativna, mišljenje je negativno. U trećem slučaju je neutralno. Iako postoje i drugi metodi agregiranja, ovaj jednostavni algoritam daje dobre rezultate u mnogim slučajevima.

### 3.3.4.1. Sumiranje mišljenja

Posmatranje pojedinačnih mišljenja potrošača nije dovoljno usled subjektivnosti mišljenja. Većina aplikacija sentiment analize zahteva izučavanje mišljenja velikog broja korisnika i neku vrstu njihovog sumiranja. Ovakvo sumiranje mišljenja nazvano je sumiranjem baziranim na aspektima, odnosno na karakteristikama. Ono je usmereno na entitete, aspekte i mišljenja o njima, a poseduje i kvantitativnu stranu, koja je esencijalna za aspektno orijentisano sumiranje mišljenja. Kvantitativni aspekt sumiranja podrazumeva pružanje broja ili procenta osoba koje imaju pozitivan ili negativan stav prema određenom entitetu ili aspektu. Autori radova (Hu, et al., 2004) i (Liu, et al., 2005) dali su radni okvir na koji se brojna istraživanja naslanjaju i koji se široko primenjuje u industriji. Ukoliko se mišljenje definiše kao petorka  $(e_i, a_{ij}, s_{ijkl}, h_k, t_1)$ , gde je  $e_i$  naziv entiteta,  $a_{ij}$  aspekt entiteta  $e_i$ ,  $s_{ijkl}$  sentiment o aspektu  $a_{ij}$  entiteta  $e_i$ ,  $h_k$  nosilac mišljenja, a  $t_1$  je vreme kada je  $h_k$  iskazao mišljenje, ono daje dobar izvor informacija i potreban okvir za generisanje kvantitativnih i kvalitativnih rezimea. Autori su dali primer sumiranja mišljenja o foto-aparatima. Rezultat sumiranja je dat u strukturiranoj formi, kako ilustruje primer 1. *Opšti aspekt* daje prikaz uopštenog mišljenja potrošača o foto-aparatu, odnosno o entitetu kao celini, gde se vidi da 105 potrošača ima pozitivno mišljenje, dok 12 ima negativno. *Kvalitet fotografije* i *Trajanje baterije* predstavljaju dva aspekta foto-aparata. Devedeset pet potrošača ima pozitivno mišljenje o kvalitetu fotografije, 10 negativno, dok 50 ima pozitivan stav o dužini trajanja baterije, a 9 negativan. <Pojedinačne rečenice recenzije> pružaju link ka individualnim

rečenicama ili celim komentarima u kojima su iskazana mišljenja. Uz ovakav način sumiranja mišljenja sa lakoćom se može identifikovati sveukupan stav o entitetu, kao i o pojedinačnim aspektima uz mogućnost dekompozicije (engl. *drill-down*) na pojedinačne recenzije. Strukturiran rezime mišljenja pruža mogućnost vizualizacije mišljenja o pojedinačnim entitetima i njegovim aspektima, kao i paralelni prikaz mišljenja o različitim entitetima (Liu, et al., 2005).

*Foto-aparat 1:*

Aspekt: Opšti	
Pozitivno: 105	<Pojedinačne rečenice recenzija>
Negativno: 12	<Pojedinačne rečenice recenzija>
Aspekt: Kvalitet fotografije	
Pozitivno: 95	<Pojedinačne rečenice recenzija>
Negativno: 10	<Pojedinačne rečenice recenzija>
Aspekt: Trajanje baterije	
Pozitivno: 50	<Pojedinačne rečenice recenzija>
Negativno: 9	<Pojedinačne rečenice recenzija>

...

**Primer 1. Sumiranje mišljenja zasnovano na aspektima**

Mišljenja data u formi petorke pružaju dalje mogućnosti strukturiranja mišljenja. Ukoliko je ekstrahovano vreme može se prikazati trend mišljenja o različitim aspektima. Čak i bez analize mišljenja, mogu se prikazati frekvencije pominjanja svakog aspekta koje će dati sliku o dominantnom aspektu u datom vremenu (Liu, 2012). Nad petorkama mišljenja mogu se primeniti brojne OLAP<sup>28</sup> tehnike kako bi se podaci bolje shvatili (npr. tehnike *slice* i *dice*). Pored poređenja mišljenja korisnika po različitim entitetima, rezultati sentiment analize se mogu koristiti kao „sirovi“ podaci za *data mining*. Tako npr. primena tehnike klasterovanja može dati interesantne tržišne segmente i uvid u njihove stavove.

Sumiranje mišljenja, međutim, ne mora rezultovati ovako strukturiranim rezimeom. Može biti i u formi teksta gde se često ne uzimaju u obzir pojedinačni aspekti i nema kvantitativnu komponentu.

---

<sup>28</sup> OLAP (engl. *Online Analytical Processing*) se odnosi na tehnike koje se koriste u poslovnoj inteligenciji za otkrivanje podataka, izveštavanje, kompleksne analize podataka i predviđanje.

## 4. PRIMENA SENTIMENT ANALIZE U VISOKOM OBRAZOVANJU

Sentiment analiza srpskog jezika predstavlja nedovoljno istraženu oblast. Iako su prisutni sporadični pokušaji u domenu analize sentimenta ((Milošević, 2013), (Analiza emocija, Matematički fakultet u Beogradu, 2011)), odnosno u domenu obrade prirodnog jezika ((Korpus savremenog srpskog jezika, 2002), (Vebranka, 2007), (Ševa, et al., 2003), (Sečujski, et al., 2006)), osnovnu prepreku u razvoju sentiment analize čine nedovoljni, u nekim aspektima čak nepostojeći resursi, neophodni za njenu realizaciju, a prevashodno nedostatak anotiranih korpusa za različite domene. Usled izazova sa kojima se visoko obrazovanje u Srbiji susreće, poput privlačenja novih i zadržavanja postojećih studenata, kao i problema koji se odnose na nekvalitetnu evaluaciju nastavnog osoblja i institucije kao celine, centralni deo doktorske teze predstavlja manuelna anotacija korpusa za domen visokog obrazovanja, koja podrazumeva dodavanje semantičkih i drugih interpretativnih informacija kolekciji tekstova. Anotacija korpusa je dragocena pre svega jer se njom stvara osnovna baza znanja za obučavanje algoritama mašinskog učenja za sprovođenje sentiment analize, kao i za validaciju teorija o pojavama u tekstovima (strukturnih, logičkih, semantičkih i sintaksičkih) (Hovy, et al., 2010). Anotacija korpusa za domen visokog obrazovanja stvorice osnovu i za prevazilaženje gorepomenutih izazova.

Kao što je istaknuto u [Poglavlju 1.3](#), privlačenje studenata predstavlja jedno od ključnih pitanja rada i opstanka visokoškolskih institucija u Republici Srbiji zbog veoma izraženog demografskog pada, koji je doveo do fenomena poznatog kao „bela kuga“. Do 2020. godine značajan pad nataliteta u Srbiji će rezultovati daleko manjim brojem svršenih srednjoškolaca od ukupnog broja svih raspoloživih mesta za prijem na visokoškolske ustanove<sup>29</sup>, što znači da predstoji „borba“ za svakog studenta, i samo one institucije koje uspeju da privuku zadovoljavajući broj studenata moći će uspešno da posluju. Primenom sentiment analize za praćenje zadovoljstva studenata u realnom vremenu, identifikovanje uskih grla u poslovanju (slabosti predmeta i predavača, identifikovanje problema studenata na predmetima gde je mala prolaznost, identifikovanje razloga zbog kojih studenti napuštaju visokoškolsku instituciju, organizacioni propusti institucije, problemi sa pojedinim službama fakulteta i sl.), kao i za praćenje tržišnih trendova, visokoškolske institucije mogu steći neophodno znanje za preduzimanje korektivnih mera, za negovanje reputacije i prepoznatljivosti brenda, gradnju i jačanje lojalnosti postojećih studenata, kao i prilagođavanje usluga preferencijama studenata i tržišnim trendovima. Viši stepen zadovoljstva postojećih studenata donosi pozitivan marketing od usta do usta koji ima značajan uticaj na privlačenje novih studenata ((Barreda, et al., 2015), (Chow, et al., 2015), (Phillips, et al., 2015)).

---

<sup>29</sup> Strategija razvoja obrazovanja u Srbiji do 2020. godine je dostupna na sajtu: [http://www.srbija.gov.rs/vesti/dokumenti\\_sekcija.php?id=45678](http://www.srbija.gov.rs/vesti/dokumenti_sekcija.php?id=45678)

Prema Zakonu o visokom obrazovanju Republike Srbije, član 15<sup>30</sup>, visokoškolske institucije su u obavezi da periodično sprovode evaluaciju kvaliteta ustanove, studijskih programa, nastave i uslova rada. Ovakva evaluacija uključuje i obaveznu ocenu studenata. Na Univerzitetu u Novom Sadu studentska evaluacija se realizuje kroz strukturirane upitnike<sup>31</sup> i predstavlja jedinu formalnu povratnu spregu sa studentima. Popunjeni upitnici mogu otkriti korisne informacije, poput toga da li se nastava odvija redovno i po rasporedu, stepen razumljivosti predavanja, kvalitet nastave (studenti evaluiraju ove aspekte ocenama od 5 do 10). Međutim, ovakve informacije najčešće nisu dovoljno informativne kako bi se upotrebile za unapređenje poslovanja ili lično napredovanje nastavnog osoblja, jer ukazuju na sveukupno zadovoljstvo, odnosno nezadovoljstvo, dok znanje o specifičnim karakteristikama profesora, stilu predavanja ili odnosu profesora prema studentima, kao i problemima na pojedinim predmetima, uglavnom ostaje skriveno. Evaluacija kroz strukturirane upitnike takođe povlači pitanje njene objektivnosti i kvaliteta. Upitnici se najčešće fizički distribuiraju studentima neposredno pre ispita, kada su studenti dekoncentrisani, ili kroz onlajn sistem kao preduslov za prijavu ispita, kada student u strahu da će se evaluacija povezati sa njihovim imenom, dodeljuju profesorima više ocene, čime se gubi smisao evaluacije. Kratkim intervjuom studenata treće godine osnovnih studija i prve godine master studija Ekonomskog fakulteta u Subotici ustanovljeno je da su retki studenti koji se potrudu da kvalitetno popune upitnike. Većina na ovakav vid evaluacije gleda vrlo kritično, smatrajući da se na ovaj način ne može postići kvalitetna evaluacija profesora i predmeta. Usled brojnosti i dužine pitanja studenti ih ne čitaju pažljivo, a kako se rezultati evaluacije ne objavljuju, gube interesovanje za ovaj vid evaluacije i do kraja studija popunjavaju upitnik potpuno nasumično i prema trenutnom raspoloženju. Potrebno je napomenuti da visokoškolske institucije mogu samostalno sprovoditi različite studentske ankete u cilju obezbeđenja kvaliteta, ali one nisu standardizovane, za razliku od strukturiranih upitnika koje propisuju univerziteti za sve svoje članice, i sprovode se sporadično.

Forumi i sajtovi za recenziranje i ocenjivanje profesora<sup>32</sup> pružaju studentima platformu za slobodnije iskazivanje mišljenja i stavova u momentu koji njima odgovara, bez pritiska institucija i profesora. Sadržaji dostupni na ovim sajtovima otkrivaju emocije studenata, njihova oduševljenja i kritike, kao i konkretne razloge za njihova (ne)zadovoljstva, kao što je ilustrovano narednim primerima preuzetim sa sajta *Oceni profesora*<sup>33</sup>:

---

<sup>30</sup> Zakon o visokom obrazovanju je dostupan na sajtu: <http://mpn.gov.rs/propisi/zakoni/obrazovanje-i-vaspitanje/505-zakon-o-visokom-obrazovanju>.

<sup>31</sup> Strukturirani upitnici imaju predefinisane formu pitanja i ponudene odgovore od kojih ispitanik odabira jedan ili više odgovora ili ih ocenjuje na zadatoj skali vrednosti.

<sup>32</sup> Rate my professor: <http://www.ratemyprofessors.com>.

Rate my teacher: <http://www.ratemyteachers.com/>.

Ehnezamo: <http://www.ehnezamo.com/>.

Ocijeni profesora: <http://www.srednja.hr/ocijeni-profesora>.

Studokracija: <http://studokracija.com/>.

MeinProf.org: <http://www.meinprof.org/>.

<sup>33</sup> Sajt *Oceni profesora* dostupan je na sledećoj URL adresi: [www.oceniprofesora.com](http://www.oceniprofesora.com).

- a) Deluje *hladno* i *profesionalno* u odnosu prema studentima, ali je u stvari jedna *vrlo fleksibilna* i *lagana* osoba. *Zna* svoj posao i ograničenja ljudi kojima predaje, što je *retkost*. *Draga* osoba, kao što je neko napisao.
- b) *Izuzetan* čovek i profesor. *Pametan*, *školovan* i *stručan* čovek. Materija po prirodi *nije baš živa*, ali je profesor iznosi na *najbolji* mogući način da je *razumeju* studenti. *Izuzetna* komunikacija sa profesorom, kao i *razumevanje* koje *cenjeni* profesor iskazuje prema studentima... Sve reči *hvale*!
- c) Obrazložiću svoje glasanje temeljno. Predavanja *nisu korisna* ni u kom kontekstu. Predavanja treba da se *PREDAJU*, a *ne da se čita* sa slajda sopstvena skripta. Samim tim, druga tačka, odnosno stepen interesantnosti predavanja, *jednak je nuli*. *Nije zanimljivo* čitati sa slajdova puke informacije. Što se tiče komunikacije, ona je *u redu*, mada ni od toga *nema koristi*, jer se opet sve svede na početnu tezu. Materijal za učenje jeste *dostupan*, ali je *nerazumljiv*, *nabacan*, *bez reda*, *bez* nekog *razumljivog rečnika*. Takođe, moram da dodam da se rešenja testova koje profesorka daje studentima *ne poklapaju* sa informacijama iz njene skripte. Prema tome, ocena 1.75 je zaista zaslužena.
- d) Što se tiče predavanja mislim da su *ok* i ima šta *zanimljivo* da se nauči. Ali sve ostalo je *veoma loše*, od kvaliteta knjige i *nesvrshodnosti* tekstova u njoj, koje je preuzela sa vikipedije, što je veoma *neprofesionalno*, pa sve do samog ispita, koji je više nego *pristrasan* (da ne upotrebim neku jaču reč). Većina studenata smatra da je *veoma loš* profesor (niko ništa novo *nije naučio* na fakultetu), a pogotovo je *loš* pedagog. Ako su ovo osobine koje profesor treba da ima onda bi sama uprava fakulteta trebala da se *zapita* koliko su joj kvalitetni kadrovi, a samim tim i kvalitet obrazovanja koji nudi i na koji je veoma ponosan.

Sentiment analizom korisnički generisanih sadržaja može se identifikovati generalna naklonost javnosti prema instituciji, što univerziteti mogu upotrebiti da bi istražili uticaj javnog raspoloženja na izbor fakulteta ili studijskih programa; mogu se agregirati dobre i loše karakteristike profesora i predmeta, kao i sentiment prema profesoru, predmetu i drugim aspektima poslovanja visokoškolskih institucija, i upotrebiti direktno za unapređenje servisa visokoškolskih institucija.

Kao što je gorepomenuto, za većinu domena u Srbiji nedostaju neophodni resursi za realizaciju sentiment analize, samim tim i za domen visokog obrazovanja. Da bi algoritmi nadgledanog učenja mogli kvalitetno da se primene za procesiranje prirodnog jezika i sentiment analizu, potrebno je obezbediti nekoliko hiljada primera označenog teksta za obučavanje, koji će pored samog teksta sadržati i dodatno, dobro definisano znanje. Konsekventno, fokus sprovedenog istraživanja je kreiranje reprezentativnog skupa podataka – korpusa – primenom postupka anotacije. Ovim će se omogućiti analiza korpusa na osnovu seta specifičnih atributa dobijenih u postupku anotacije.

Anotacija korpusa ili označavanje može se konceptualizovati kao proces obogaćivanja korpusa (dokumenata, rečenica ili reči) dodavanjem različitih lingvističkih meta podataka ili drugih

informacija od interesa u službi teorijskog ili praktičnog cilja ((Hovy, et al., 2010), (Wissler, et al., 2014)). Informacije koje će se korpusu pridružiti zavise od svrhe anotacije, a mogu varirati od sintaksičkih preko leksičkih do semantičkih informacija. Anotacija korpusa se može realizovati manuelno, kada informacije pridružuju ljudi, ili mašinski, kada se programira automatizovan sistem za označavanje. Oba pristupa anotaciji imaju prednosti i nedostatke. Automatizovana anotacija je brža i može za kratko vreme dati rezultate nad velikim korpusom. Međutim, ona zahteva značajnu pripremu korpusa i vreme potrebno za programiranje automatizovanog sistema za označavanje, a rezultati mogu biti lošeg kvaliteta. Kako je istaknuto u (Hovy, et al., 2010), relativno zadovoljavajuć stepen tačnosti je postignut automatizovanim označavanjem delova rečenice (*engl. Part-of-Speech tagging – POS*) za engleski jezik. Međutim, zadaci poput označavanja semantičkih relacija i strukture diskursa kompleksne su za automatizaciju. Ovi zadaci zahtevaju prvo manuelnu anotaciju korpusa kako bi se dobio manji korpus nad kojim se algoritam mašinskog učenja može obući, što se jedino može postići kroz dobro osmišljenu i pouzdanu proceduru anotacije. Ručna anotacija, iako od izuzetne važnosti, takođe ima svoje nedostatke. Pre svega, ona je vremenski zahtevna i skupa u smislu pronalaženja ljudi eksperata – anotatora, njihove obuke i nadzora tokom procesa anotacije (Hahn, et al., 2008), ali ne zahteva obimnu pripremu korpusa za sam postupak anotacije.

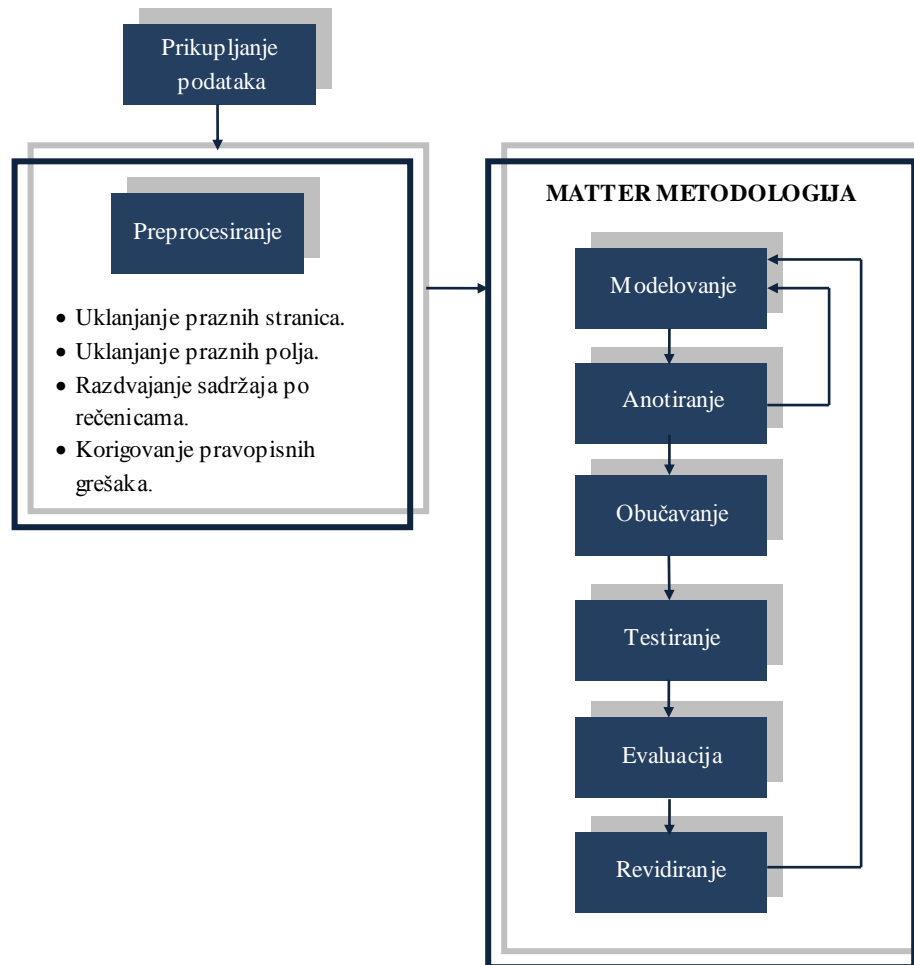
Naredno poglavlje detaljno opisuje metodologiju istraživanja u kojem je realizovana ručna anotacija korpusa, kojeg čine sadržaji preuzeti sa sajta za recenziranje nastavnog osoblja u Republici Srbiji – *Oceni profesora*, i konsekvantna sentiment analiza ovako pripremljenog sadržaja.

#### 4.1. Metodološki koraci istraživanja

U realizaciji sentiment analize praćena je *MATTER* (*engl. Model, Annotate, Test, Train, Evaluate, Revise*) metodologija prikazana u radovima ((Pustejovsky & Stubbs, 2012), (Pustejovsky & Moszkowicz, 2012), (Stubbs, 2012)). Kako su autori istakli, atributi upotrebljeni za kodiranje specifičnog jezičkog fenomena moraju biti dovoljno bogati da bi se u algoritmu koji se obučava zadržalo i oslikalo željeno ponašanje ili patern. Lingvistički opisi formiraju osnovu za anotirane vrednosti, kao i attribute koji se koriste u razvojnom ciklusu za obučavanje i testiranje algoritma čiji će zadatak biti identifikovanje ili postavljanje oznake nad tekstom. Na osnovu analize i evaluacije performansi sistema razvijeni model fenomena se može revidirati i modifikovati kako bi se ponovo pristupilo obučavanju i testiranju algoritma. Ovakav razvojni ciklus autori su nazvali *MATTER* metodologijom.

Predložena metodologija obuhvata šest koraka. *Modelovanje* podrazumeva generisanje strukturnih opisa koji proizilaze iz empirijskih zapažanja nad podacima, a koji će dati attribute. Tokom *anotacije* definiše se anotaciona shema. Ona obuhvata set definisanih atributa koji kodiraju specifične strukturne opise i svojstva input podataka. Naredni koraci podrazumevaju

upotrebu korpusa anotiranog setom definisanih atributa za **obučavanje** algoritma i njegovo **testiranje** delom korpusa koji nije upotrebljen u fazi obučavanja, nakon čega sledi **evaluacija** rezultata. Model i specifikacija anotacije se **revidiraju** kako bi se anotacija učinila robusnijom i pouzdanijom. Da bi se uspešno modelovalo sprovedeno istraživanje, predložena metodologija je proširena kako bi obuhvatila dva, pre svega, vremenski zahtevna koraka: prikupljanje i preprocesiranje tekstualnih sadržaja. Ovako prošireni metodološki koraci prikazani su na slici 4.1.



Slika 4.1. Proširena MATTER metodologija

U [Poglavlju 4.1.1.](#) opisani su postupci prikupljanja i preprocesiranja podataka. Modelovanje sadržaja za potrebe anotacije opisano je u [Poglavlju 4.1.2.](#) Pregled i opis atributa anotacije dat je u [Poglavlju 4.1.3.](#) MATTER ciklus nalaže višefaznu realizaciju anotacije, te su u [Poglavlju 4.1.3.1.](#) prikazane anotaciona procedura, zapažanja i podešavanja koja su se morala načiniti u fazi anotacije. Evaluacija kvaliteta anotacije je realizovana primenom *Fleiss kappa* statistike (([Fleiss, 1971](#)), ([Fleiss, et al., 2003](#))) nad uzorkom anotiranog korpusa, a rezultati su prikazani u [Poglavlju 4.1.3.2.](#) Distribucija tagova u korpusu i detaljan pregled zaključaka o stilu pisanja, koji

se generalno može povezati sa srpskim jezikom, lingvističkim opservacijama i pravilnostima povezanim sa sentimentima i aspektima identifikovanim u anotiranom sadržaju opisani su u [Poglavlju 4.1.3.3](#). U [Poglavlju 4.1.4](#). opisan je postupak formiranja rečnika sentiment pojmova upotrebljenih u sentiment analizi. Treći korak MATTER metodologije – obučavanje – prikazan je u [Poglavlju 4.1.5](#), dok su testiranje, evaluacija i revidiranje opisani u [Poglavlju 4.1.6](#).

#### **4.1.1. Prikupljanje i preprocesiranje podataka za osnovni skup iz društvenih medija**

Korpus predstavlja skup podataka, odnosno kolekciju tekstova koja se koristi za lingvističku analizu. Sačinjen je od reči koje nisu osmišljene od strane lingvista, nego predstavljaju autentične reči koje se prirodno pojavljuju u pisanom ili govornom jeziku. Sa aspekta korpusne lingvistike, korpus mora zadovoljiti dva osnovna uslova, mora biti reprezentativan i balansiran (Pustejovsky, et al., 2012). Kako on predstavlja podskup reči odabranog jezika, ne može obuhvatiti sve primere upotrebe reči u jeziku. Iz tog razloga se korpus kreira uzorkovanjem postojećih tekstova, pri čemu je potrebno voditi računa o reprezentativnosti prikupljenih podataka kako bi obuhvaćeni primeri oslikali punu varijabilnost populacije. Populacija koja se uzorkuje određena je ciljem analize i ciljem anotacije. Osnovni cilj istraživanja, odnosno analize, može se definisati kao sprovođenje sentiment analize tekstova pisanih na srpskom jeziku u kojima se pominje visoko obrazovanje, dok je osnovni cilj anotacije da se u recenzijama studenata identifikuje polaritet iskazanog sentimenta prema aspektu o kojem student piše. Imajući ovo u vidu, korpus koji će predstavljati osnovu za istraživanje mora obuhvatiti primere različitih aspekata poslovanja visokoškolske institucije, kao i iskazanog sentimenta. Drugi uslov koji korpus treba da zadovolji odnosi se na balansiranost. Prilikom definisanja proporcije različitih vrsta tekstova, koje će biti uključene u korpus, potrebno je voditi se informisanim i intuitivnim prosuđivanjem (Sinclair, 2005). Balansiranost korpusa je teško zadovoljiti, prvenstveno kod korpusa u kojima je potrebno obuhvatiti različite tipove tekstova, na primer kada korpus oslikava jedan jezik i kada je potrebno uključiti sve tipove pisanog (novinski članci, knjige, recenzije i dr.) i govornog jezika (transkripte različitih govora). Kako je za istraživanje koje je predmet doktorske disertacije potrebno obuhvatiti komentare koji se odnose na visoko obrazovanje, uzorak će se lakše balansirati. Skupom je potrebno obuhvatiti komentare u kojima su iskazani pozitivni i negativni stavovi o definisanim aspektima poslovanja visokoškolskih institucija. Aspekti su detaljnije opisani u [Prilogu D.1](#).

Koncepti reprezentativnosti i balansiranosti se ne mogu lako definisati. Pitanje o kojem se još uvek raspravlja u korpusnoj lingvistici jeste da li se za ma koji korpus može reći da je zadovoljio navedene preduslove. Ipak, tokom kreiranja korpusa treba težiti ka njihovom maksimalnom zadovoljenju u skladu sa postavljenim ciljevima analize i anotacije ((Pustejovsky, et al., 2012), (Sinclair, 2005)).



Osnovni izvor podataka za sprovedeno istraživanje predstavljaju javno dostupni sadržaji na portalu *Oceni profesora*, koji omogućava studentima da dodele ocene profesorima i da recenziraju njihov rad. Na sajtu su obuhvaćeni svi univerziteti u Srbiji, privatni i državni, kao i celokupno nastavno osoblje. Sajt je pušten u funkciju polovinom 2012. godine. Marta 2013. godine je imao preko 100.000 posetilaca i 930.000 pregledanih stranica, dok je marta 2014. godine taj broj porastao na 1.5 milion poseta<sup>34</sup>. Za 2015. godinu broj ocena do juna meseca dostigao je 44.000.

Sajt funkcioniše po principima društvenih mreža:

1. U osnovi je **koncept zajednice**. Kako se i zajednice ili društvene grupe u realnom svetu formiraju na osnovama zajedničkog interesovanja ili uverenja, i onlajn društvene mreže su bazirane na ovom principu. Portal *Oceni profesora* kroz kreiranje korisničkih profila stvara zajednicu studenata određenog fakulteta, koji razmenjuju mišljenja o profesorima i dele svoja iskustva sa članovima te zajednice.
2. Zasnovan je na **korisničkim profilima**. Pre pojave sajtova društvenih mreža i društvenih medija, sajtovi su bili bazirani na sadržaju koji je kreiran od strane vlasnika sajta, a posetioci sajta su imali samo mogućnost čitanja prezentovanog sadržaja. Tok informacija je bio jednosmeran, a ažuriranje sadržaja je vršio vlasnik sajta ili administrator. Sa druge strane, sadržaji sajtova koji imaju karakteristike društvenih mreža stvaraju se dinamički i njima upravljaju sami korisnici sajta. Bez korisnika sajt ne bi imao sadržaj, a tok informacija određuje svaki korisnik koji uzme učešće na ovom sajtu.

Kreiranjem naloga/profila korisnici sajta *Oceni profesora* se pridružuju podzajednici fakulteta koji pohađaju i imaju mogućnost recenziranja profesora koji na tom fakultetu predaju. Dodatno, profesori mogu kreirati svoje naloge kako bi samostalno vodili svoju stranicu, ažurirali podatke i odgovarali na studentske recenzije.

3. **Kreiranje stranica**. Ukoliko nedostaju stranice o profesorima, članovima zajednice data je mogućnost proširenja sadržaja dodavanjem novih stranica i sadržaja o instituciji i predmetima na kojima je profesor angažovan, kontakt podataka i drugih informacija koje student smatra relevantnim.

Portal ima jasno definisan opseg, recenziranje nastavnog osoblja, što je omogućilo implementaciju ocenjivanja pojedinih domenski specifičnih **aspekata**, slika 4.2. Dodeljivanjem 1 do 5 zvezdica studenti evaluiraju sledeće aspekte: korisnost predavanja, interesantnost predavanja, korektnost komunikacije i dostupnost i kvalitet materijala. Sajt koristi pojedinačne ocene kako bi kalkulisao sveukupnu ocenu profesora. **Tekst recenzije** je polje u kojem studenti iskazuju svoje misli i stavove prema profesoru kojeg ocenjuju. Ovo je opciono polje i nema

---

<sup>34</sup> Podaci su preuzeti sa zvanične facebook stranice sajta *Oceni profesora*, koja je dostupna na sledećoj URL adresi: <https://www.facebook.com/rangiranjeprofesora>.

ograničenje po pitanju dužine teksta. Kako je istaknuto na zvaničnoj *facebook* stranici sajta, svaka deveta ocena sadrži i tekst recenzije<sup>35</sup>. Studenti se podstiču da iskažu slaganje ili neslaganje sa onim što je navedeno u tekstu recenzije odabirom dugmeta „*sviđa mi se*“, odnosno „*ne sviđa mi se*“, što odražava **kvalitet komentara**. Svakom tekstu recenzije su dodeljeni **datum i vreme** kada je recenzija napisana.



Slika 4.2. Prikaz stranice profesora na sajtu Oceni profesora

Za preuzimanje javno dostupnih sadržaja sa sajta *Oceni profesora* upotrebljen je skript napisan u *Python* programskom jeziku, baziran na algoritmu prikazanom u [Prilogu A](#). Sa portala je preuzeto 2284 stranice, gde svaka stranica predstavlja profil jednog profesora za kojeg je barem jedan student ostavio komentar, dok je najviše pet komentara preuzeto za jednog profesora. Stanice su sačuvane u tekstualnom (.txt) formatu. Jedna preuzeta stranica predstavlja profil jednog profesora sa osnovnim podacima o profesoru i pridruženim recenzijama. Primer sadržaja jedne preuzete stranice:

['Otto Barak , PhD', 'Medicinski fakultet', 'Univerzitet u Novom Sadu', 'Kratka biografija', 'Curriculum vitae', '04.10.2012. | 23:21:47', 'Odlican i drugaciji nacin predavanja od vecine ostalih profesora. Neka tako nastavi.', ' (0)', ' (0)']

<sup>35</sup> Podaci su dostupni na zvaničnoj facebook stranici sajta Oceni profesora: <https://www.facebook.com/rangiranjeprofesora/posts/615880361892094>

Nakon ručnog preprocesiranja sadržaja, koje je pored uklanjanja praznih stranica i stranica bez komentara podrazumevalo razdvajanje i poravnanje sadržaja po kolonama, dobijen je dokument u kojem jedan red opisuje jednog profesora sledećim atributima:

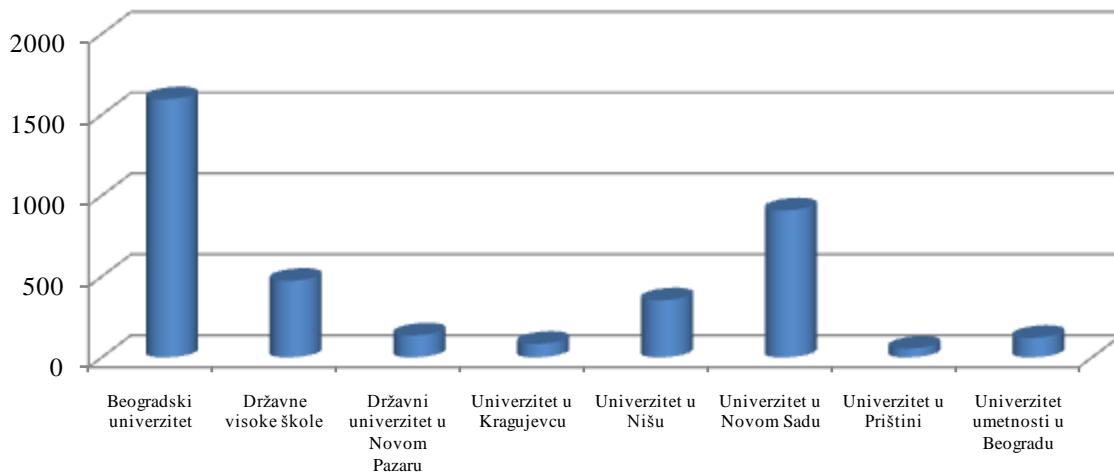
- Lokacija sa koje je sadržaj preuzet
- Ime i prezime profesora
- Fakultet na kojem profesor predaje
- Univerzitet kojem fakultet pripada
- Datum kada je komentar postavljen
- Komentar
- Ocene studenta koji je ostavio komentar o korisnosti predmeta, stepenu u kojem su predavanja (ne)interesantna, korektnosti komunikacije i dostupnosti i kvalitetu materijala
- Koliko puta su različiti studenti odabrali dugme *Like* (svidi mi se) za dati komentar
- Koliko puta su različiti studenti odabrali dugme *Dislike* (ne svidi mi se) za dati komentar
- Identifikaciona oznaka komentara (manuelno pridružena oznaka).

Jedan profesor se može pojaviti maksimalno pet puta u skupu podataka, s obzirom na to da je najviše pet komentara preuzeto za svakog profesora. Tabela 4.1. ilustruje strukturu dokumenta nakon navedenog preprocesiranja podataka, bez izvora i imena profesora.

**Tabela 4.1. Izgled dokumenta nakon preprocesiranja**

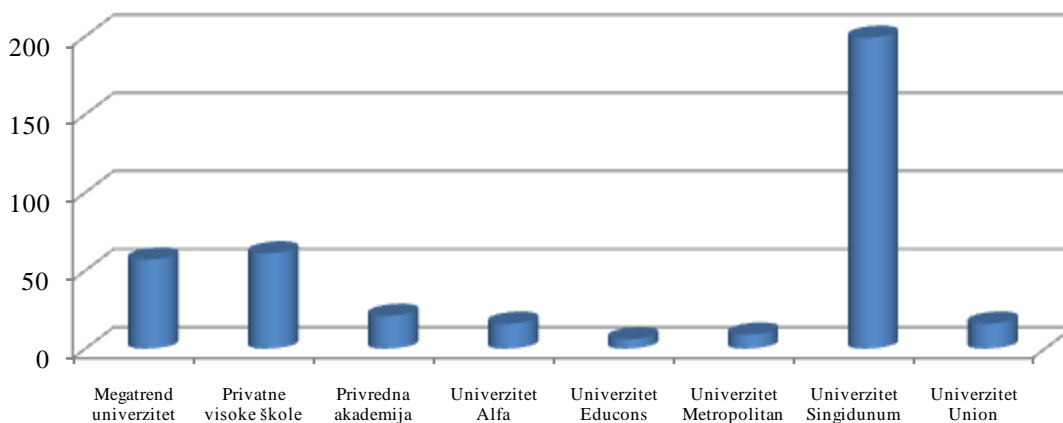
Fakultet	Univerzitet	Datum	Komentar	P.Korisna	P.Interesantna	Komunikacija	Materijali	Like	Dislike	Oznaka
Filozofski fakultet u Nišu	Univerzitet u Nišu	09.01.2013.   17:46:09	Da nam je vise takvih profesora sve bi bilo sjano...Sve pohvale.....	5	5	5	5	23	21	260
Medicinski fakultet	Beogradski Univerzitet	13.03.2014.   23:58:11	Sjajna je ,odlican pedagog ,sve pohvale	5	5	5	5	4	0	530

Ukupno je preuzeto 4159 komentara. Slika 4.3. i slika 4.4. prikazuju njihovu distribuciju po univerzitetima. Kao što je ilustrovano na slici 4.3, shodno veličini univerziteta i broj dostupnih i preuzetih komentara se uvećava. Tako u preuzetim sadržajima dominiraju recenzije profesora koji predaju na Beogradskom Univerzitetu i Univerzitetu u Novom Sadu.



Slika 4.3. Distribucija komentara po državnim univerzitetima

U kreiranju sadržaja na portalu *Oceni profesora* najaktivniji su studenti Univerziteta Singidunum, slika 4.4. Ovakva situacija je razumljiva s obzirom na to da su inicijatori i kreatori ovog portala studenti Fakulteta za informatiku i računarstvo Univerziteta Singidunum, kojima su se kasnije priključili studenti Fakulteta organizacionih nauka, Matematičkog fakulteta i Pravnog fakulteta Univerziteta u Beogradu.



Slika 4.4. Distribucija komentara po privatnim univerzitetima

Kako je predmet istraživanja sentiment analiza sadržaja na srpskom jeziku, komentari koji su pisani na engleskom, španskom i latinskom jeziku uklonjeni su iz korpusa. Takođe, predmet anotacije su izjavne, uzvične i željne rečenice, jer se u njima može prepoznati sentiment, te su iz korpusa uklonjene upitne rečenice koje samostalno čine komentar. Uklonjeno je 285 takvih komentara, što čini konačni korpus od 3863 komentara.

Skript za preuzimanje sadržaja ne podržava UTF-8 kodiranje teksta. U preuzetim sadržajima slova sa dijakritičkim znacima (š, đ, ž, č, ć) bila su automatski zamenjena odgo varajućim kodom, te je bilo potrebno izvršiti njihovu ponovnu konverziju u slova sa dijakritičkim znacima.

Kao većina korisnički generisanog sadržaja, i preuzeti komentari su uglavnom pisani neformalnim stilom, bez poštovanja pravopisnih pravila (izostanak praznog mesta nakon interpunkcijskih znakova, upotreba velikog broja uzvičnika i tačaka na kraju rečenice, neretko i na polovini rečenice, izostanak interpunkcijskog znaka na kraju rečenice, i brojni drugi propusti), bez dijakritičkih znakova i bez poznavanja osnovnih gramatičkih pravila, što je dodatno otežalo pripremu podataka. Kako bi se uklonile pravopisne greške, upotrebljen je onlajn alat Hascheck<sup>36</sup>.

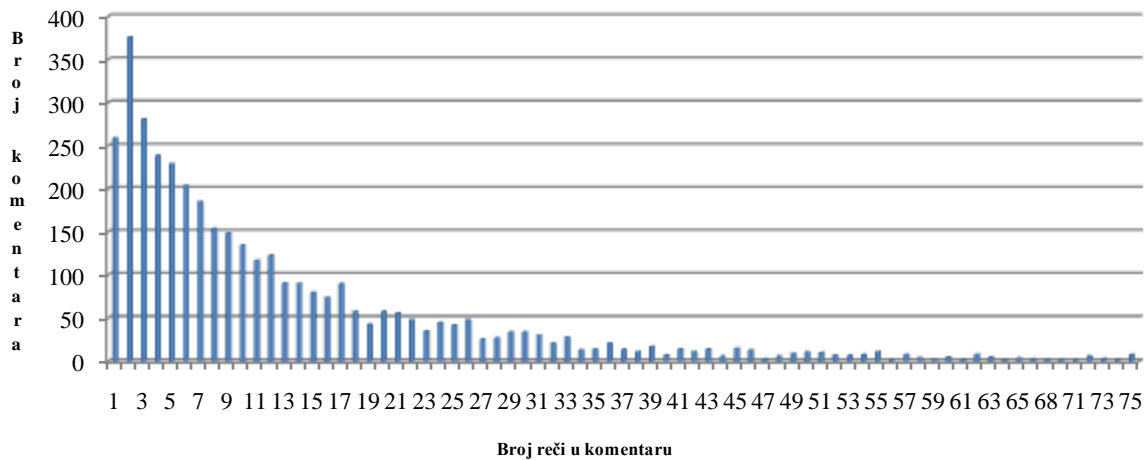
Na osnovu prikupljenih podataka kreiran je ručno anotirani korpus. Manuelna anotacija je realizovana na osnovu anotacione sheme prikazane u narednom poglavlju. Anotacija je bazirana na aspektima i vrši se na nivou rečenice ili dela rečenice. Iz ovog razloga izvršen je naredni korak preprocesiranja podataka, koji je podrazumevao podelu komentara na rečenice. Iako su javno dostupni alati koji uspešno realizuju ovaj zadatak pre svega za engleski jezik, potom za druge jezike koji su široko rasprostranjeni, poput nemačkog, kineskog, španskog, ovi alati nisu mogli da se upotrebe nad sadržajima pisanim na srpskom jeziku. Dostupni alati prelamaju tekst nakon univerzalnih interpunkcijskih znakova koji ukazuju na prestanak važenja rečenice. Međutim, naš jezik nameće ograničenja u njihovoj primeni, koja se prevashodno odnose na skraćenice s tačkom, karakteristične isključivo za naše govorno područje, kao i na drugačiji način pisanja datuma i rednih brojeva. Neke skraćenice u srpskom jeziku pišu se sa tačkom na kraju, poput npr., pr., prof., dipl.. Postojeći alati za razdvajanje rečenica izvršili su nepravilno prelamanje rečenica u tekstovima u kojima su prisutne ove i druge skraćenice. Dodatno, prisustvo rednih brojeva, koji se u srpskom jeziku takođe pišu s tačkom, dovodi do pogrešnog prelamanja rečenica. Nadalje, upotreba višestrukih interpunkcijskih znakova, više uzastopnih uzvičnika i tački, dovodi do prelamanja rečenice nakon prvog interpunkcijskog znaka, dok bi se ostatak interpunkcijskih znakova tretirao kao zasebna rečenica. Usled ovih specifičnosti primena već razvijenih alata za razdvajanje rečenica nad prikupljenom kolekcijom tekstova nije moguća. Za potrebe istraživanja razvijen je razdelnik rečenica kojim su obuhvaćene navedene specifičnosti. Kako se ne bi remetilo razdvajanje rečenica, navedeni izuzeci se zamenjuju ekvivalentima definisanim u skriptu, potom se vrši razdvajanje rečenica. **Prilog B** ilustruje algoritam primenjen u razdvajanju rečenica, čime se izuzeci sa kojima smo se susreli u radu prevazilaze. Korpus je sačinjen iz 6896 rečenica.

Praktično pravilo za veličinu korpusa svodi se na ispitivanje postojećih korpusa koji se koriste za srodne zadatke (Pustejovsky, et al., 2012). Korpuse, predstavljene u radovima ((Hu, et al., 2004), (Ganu, et al., 2009), (Kessler, et al., 2010), (Broß, 2013), (Toprak, et al., 2010)), čini 113 do 1100 recenzija. Korpusom *Oceni profesora* obuhvaćeno je 3863 recenzije. Međutim, gorepomenute korpuse čine recenzije tehničkih proizvoda, hotelijerskih usluga ili filmova, koje

<sup>36</sup> Hrvatski akademski spelling checker <https://hcheck.tel.fer.hr/>

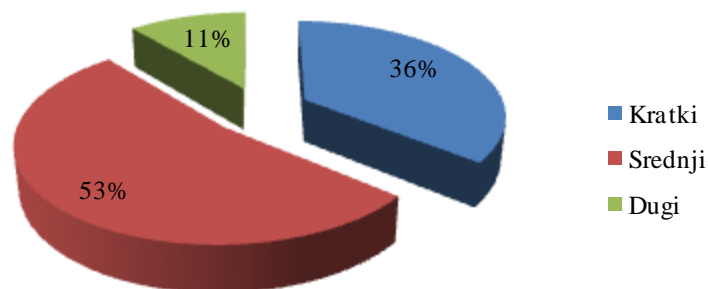
se obično pišu opširnije te je to rezultovalo korpusima koji imaju manji broj recenzija od korpusa *Oceni profesora*, a približan broj rečenica. Većina prezentovanih korpusa je sačinjena iz 2700 do 6100 rečenica, dok korpus koji je predmet istraživanja u doktorskoj tezi ima 6896 rečenica.

Dužina komentara obuhvaćenih korpusom *Oceni profesora* varira do maksimalno 180 reči. Kako slika 4.5. ilustruje, u korpusu dominiraju kratki komentari. Sa ilustracije su izostavljeni komentari koji imaju više od 75 reči jer se oni pojavljuju jednom do dva puta u korpusu. Pojedini komentari sadrže samo jednu reč, a neki čak ni toliko – u funkciji komentara se javljaju emotikoni. Pojavljuju se samostalno ili kao dopuna šireg komentara.



Slika 4.5. Pregled komentara po broju reči

Prema klasifikaciji komentara koju su dali autori u radu (Wyner, et al., 2009) kratki komentari sadrže 1-5 reči, srednji 6-30 reči, dok dugi imaju više od 30 reči. Na osnovu ovakve raspodele komentara u korpusu dominiraju komentari srednje dužine, kojih je 2047, odnosno 53%. Slika 4.6. ilustruje broj komentara po kategorijama i njihovo procentualno učešće u korpusu.



Slika 4.6. Učešće komentara u korpusu, klasifikovanih prema dužini

Ukoliko se pođe od pretpostavke da studenti koji pojedine aspekte (korisnost i interesantnost predavanja, kvalitet komunikacije i materijala) ocene višim ocenama, 4-5, ostavljaju pozitivne komentare, dok oni koji ove aspekte ocenjuju nižim ocenama, 1-2, ostavljaju negativne komentare, može se primetiti da je u korpusu prisutna blaga asimetrija u distribuciji komentara, i to u korist pozitivnih. Kako svedoče druge studije, ovakva distribucija sentimenta u korpusu nije atipična ((Cui, et al., 2006), (Dave, et al., 2003), (Gamon, et al., 2005), (Broß, 2013)). Tabela 4.2. ilustruje distribuciju komentara prema ocenama pojedinih aspekata. Može se primetiti da je distribucija ocena po aspektima ujednačena, ne postoje značajnija odstupanja. Na nivou korpusa uviđamo približno 67% pozitivnih komentara nasuprot 27% negativnih komentara.

**Tabela 4.2. Distribucija komentara prema ocenama**

Ocena	Predavanja su korisna	Predavanja su interesantna	Komunikacija je korektna	Kvalitet materijala
1	21.00%	23.98%	22.19%	20.15%
2	5.98%	5.36%	4.27%	5.31%
3	6.03%	6.91%	5.08%	7.17%
4	7.56%	7.17%	6.53%	8.80%
5	59.43%	56.58%	61.94%	58.57%

Distribucija pozitivnih i negativnih sadržaja blago se menja u korist negativnih sadržaja kada se posmatra korpus nakon anotacije, tabela 4.3. Na sajtu *Oceni profesora* ocene su pridružene jednoj recenziji. U jednoj recenziji student može da iskaže mišljenje o više različitih aspekata, otuda se anotacija vršila na nivou rečenice ili dela rečenice. Iz ovog razloga je u korpusu prisutan veći broj anotacija od broja prikupljenih recenzija, te je navedeni porast posledica promene u nivou evaluacije. Ukoliko bi se anotacija vršila na nivou recenzija, ocene, koje studenti ostavljaju uz komentare, mogle bi da se dovedu u vezu sa anotiranim recenzijama, dok u primeru konkretnog istraživanja one nisu ekvivalentne.

**Tabela 4.3. Distribucija pozitivnih i negativnih komentara u anotiranom korpusu**

	pozitivan	negativan	neutralan
% od ukupnog broja komentara	60.99%	37.66%	1.35%

Naredno poglavlje opisuje modelovanje problema na osnovu kojeg su definisane shema i postupak anotacije.

#### 4.1.2. Modelovanje sadržaja za potrebe anotacije korpusa

Kako bi se izvršilo adekvatno modelovanje fenomena koji je predmet anotacije, nasumično je izdvojeno 10% korpusa (u nastavku probni korpus) nad kojim je realizovana inicijalna eksploracija sadržaja.

Komunikacija na sajtu *Oceni profesora* je jednosmerna, studenti iskazuju svoje zadovoljstvo ili nezadovoljstvo profesorom, te su njihove poruke u formi iskaza, pozitivne ili negativne povratne informacije. Studenti ne ostvaruju dijalog sa ostalim članovima zajednice. Način na koji oni iskazuju mišljenja na sajtu *Oceni profesora* i ocene koje ostavljaju usko su povezani sa recenzijama korisnika različitih proizvoda (digitalni foto-aparati, računari, mobilni telefoni i slično), usluga (hotelijerske usluge, zdravstvene usluge), knjiga ili filmova. Istraživanja iz oblasti sentiment analizi, čiji predmet istraživanja predstavljaju recenzije potrošača, opisana su u [Poglavlju 2.1](#). Izvori podataka u prezentovanim istraživanjima srodni su izvoru *Oceni profesora*, što je omogućilo da se osnovni pristupi zauzeti u opisanim istraživanjima usvoje, uz određena prilagođavanja domenu visokog obrazovanja, jeziku i strukturi obuhvaćenih podataka. Međutim, nijedan od opisanih korpusa u [Poglavlju 2](#). ne može se direktno upotrebiti za evaluaciju anotiranog korpusa u istraživanju koje je predmet doktorske disertacije, pre svega zbog činjenice da su sadržaji i anotacije na engleskom jeziku i domeni se značajno razlikuju od onog koji je predmet doktorske teze, što povlači razlike u terminologiji.

Da bi se uspješnije sagledao i razumeo sentiment studenata skriven u recenzijama i njihov stav prema različitim aspektima rada nastavnog osoblja, osnovni zadatak istraživanja, koje je predmet doktorske disertacije, usko je povezan sa aspektno orijentisanom analizom recenzija: identifikovanjem relevantnih aspekata rada nastavnog osoblja i klasifikovanjem sentiment izraza. Istraživanja opisana u [Poglavlju 2](#). ((Broß, 2013), (Kessler, et al., 2010), (Stoyanov, et al., 2008), (Toprak, et al., 2010)) ukazuju na činjenicu da su teme, odnosno aspekti, domenski zavisni. Iz tog razloga izvršena je analiza predmeta studentskih recenzija na sajtu *Oceni profesora*. Pored ocene osnovnih aspekata (korisnost i interesantnost predavanja, korektnost komunikacije i valjanost dostupnih materijala), studenti imaju mogućnost da ostave recenziju profesora. Iako je u većini slučajeva mišljenje koje studenti iskazuju u formi recenzija usmereno na aspekte predefinisane strukturom sajta, inicijalno istraživanje sadržaja je pokazalo da oni nisu dovoljni da se adekvatno oslika struktura studentskog mišljenja. Naime, studenti pišu o načinu na koji profesori<sup>37</sup> izlažu gradivo, korisnosti predavanja, njihovom odnosu prema studentima, kriterijumu ocenjivanja, opštim karakteristikama profesora kao čoveka, pa potom kao predavača, materijalima, predmetu i organizaciji nastave. Ovakva struktura sadržaja u probnom korpusu ukazala je na potrebu za definisanjem šireg seta aspekata koji će biti predmet anotacije. Definisani aspekti su detaljnije opisani u [Poglavlju 4.1.3](#).

---

<sup>37</sup> U daljem radu se pod profesorom podrazumeva sve nastavno osoblje zaposleno na visokoškolskim institucijama: redovni profesori, vanredni profesori, docenti, asistenti sa doktoratom, asistenti, saradnici u nastavi, istraživači pripravnici.



U studentskim recenzijama identifikovano je prisustvo sentimenta, otuda je korpus neophodno obogatiti informacijama o emocijama višeg nivoa (Litman, et al., 2004), odnosno pozitivnim, negativnim i neutralnim emocijama, koje su detaljnije opisane u [Poglavlju 4.1.3.](#)

U iskazivanju emocija studenti učestalo koriste eksplicitne sentiment iskaze, odnosno reči. Da bi se dobila bogatija baza znanja i da bi se potencijalno uspešnije identifikovali pozitivni i negativni aspekti rada nastavnog osoblja, uvedena je dodatna oznaka kojom će se obeležavati sentiment iskazi. Inicijalno je vršeno označavanje pozitivnih i negativnih sentiment reči istom oznakom – *Sentiment reč*. Kako je tokom probne anotacije ovo identifikovano kao problem u smislu nedovoljne informativnosti, označavanje sentiment reči je razdvojeno na dve oznake: *Pozitivna sentiment reč* i *Negativna sentiment reč*. Sentiment iskazi/reči su detaljnije opisani i definisani u [Poglavlju 4.1.3.](#)

Pojava negacije u probnom korpusu, prevashodno u iskazivanju negativnog sentimenta, nije retka. Ukoliko se anotacijom korpusa ne bi obuhvatila negacija, iskazani sentiment se ne bi adekvatno oslikao. Otuda je identifikovana potreba za označavanjem negacije i njenog opsega nad prikupljenim sadržajem. U [Poglavlju 4.1.3.](#) detaljnije je opisan postupak njihovog označavanja.

Identifikovano je da iskazi osećanja nisu istog intenziteta. Postoje značajna odstupanja i u komentarima pohvale i pokude, što je ukazalo na potrebu za anotiranjem intenziteta iskazanog osećanja. Inicijalno je definisana skala intenziteta od 1 do 5, kojom bi se anotirao sentiment jako slabog, slabog, srednjeg, jakog i izuzetno jakog intenziteta. Na osnovu mišljenja konsultovanih eksperata iz domena psihologije<sup>38</sup> ustanovljeno je da je suviše teško napraviti jasnu distinkciju između intenziteta 4 i 5, odnosno 1 i 2, te je u skladu sa njihovim preporukama skala intenziteta sentimenta svedena na 1-3, a njome će se označavati slab, srednji i jak intenzitet iskazanog sentimenta. Eksperti su istakli da je izuzetno teško standardizovati način dodeljivanja intenziteta iskazanim stavovima s obzirom na to da oni pored samog mišljenja obuhvataju i različite emocije koje su podložne subjektivnoj interpretaciji, te da će se potencijalne poteškoće pojaviti i sa ovako svedenom skalom.

Većina korisničkih/potrošačkih recenzija iskazuje i pozitivan i negativan sentiment u jednoj recenziji, nekada čak u jednoj rečenici. U korpusu *Oceni profesora* ovo se potvrdilo kao praksa, a dodatno pored iskazivanja višestrukih polariteta sentimenta studenti u jednoj rečenici učestalo pišu i o različitim aspektima. Eksperti su se složili da je rečenicama sa dualnim aspektima najkorisnije pridružiti oba, uz pridruživanje adekvatnog polariteta sentimenta svakom. Iz ovog razloga implementiran je pristup anotaciji jezičkih pojava koji je izvan nivoa rečenice. Tako jedinica anotacije može biti deo rečenice u kojem se iskazuje jedan polaritet sentimenta i jedan aspekt. Sa ekspertima je diskutovano i u vezi sa 57 problematičnih rečenica identifikovanih u probnom korpusu. Primeri ovih rečenica navedeni su u tabeli 4.4. Analizom izdvojenih rečenica

---

<sup>38</sup> Gorana Francišković, diplomirani psiholog, zaposlena na Ekonomskom fakultetu u Subotici, i Boris Telečki, diplomirani psiholog, zaposlen u Centru za socijalni rad grada Subotice.

eksperti su razrešili određene nedoumice, pre svega u pogledu identifikovanja činjenica i određivanja iskazanog sentimenta. Naime, kada student upotrebljava izraze koji ukazuju na htenje, želju, izdvajanje pojedinca iz grupe, kada iskazuje svoje mišljenje o drugoj osobi, ovakve rečenice se ne mogu tretirati kao činjenice.

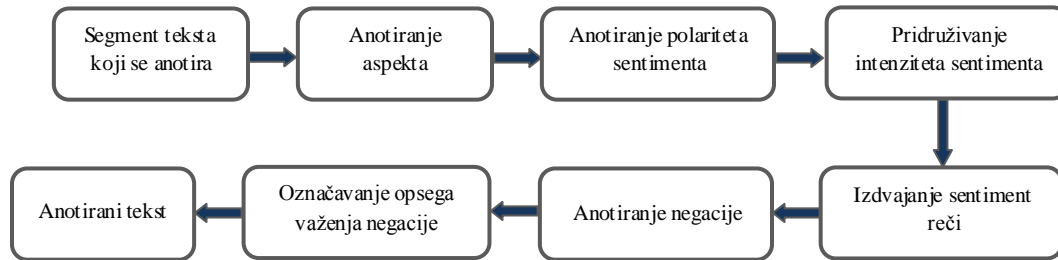
**Tabela 4.4. Primeri konfliktnih rečenica**

Rečenica	Inicijalni problem	Mišljenje eksperata
<i>Jednom je bitno odakle je Egej, a sledeći put nije bitno kako se zove Antigonin brat.</i>	Razlikovanje činjenice od subjektivnog mišljenja.	Implicitno iskazivanje nepostojanja jasnog kriterijuma pri ispitivanju.
<i>Žalba upućena na rad profesora prodekanu za nastavu je odbačena uz izgovor da se odnos zbira ocena na kolokvijumu ne računa prema pravilu koje je profesor postavio, već se podrazumeva da se KUMULATIVNO MORAJU položiti svi kolokvijumi, iako to nije naznačeno u pravilniku.</i>	Razlikovanje činjenice od subjektivnog mišljenja.	Konstatacija koja nije subjektivna.
<i>Učio od najboljih iz oblasti turizma, ali iz samo njemu znanih razloga odbio priliku da uči najbolje u oblasti turizma kada je jedini ostao na Ekonomskom fakultetu.</i>	Identifikovanje sentimenta.	Način na koji se student izražava „samo njemu znanih razloga“ ukazuje na prisutnost negativnog sentimenta.
<i>Ne želi definicije da se znaju, već mu je važnije da se gradivo razume.</i>	Razlikovanje činjenice od subjektivnog mišljenja.	Formulacija rečenice, pre svega reč „želi“, ukazuje na nečije subjektivno mišljenje, ne činjenicu.
<i>Jedna od retkih koja se javi na ulici.</i>	Razlikovanje činjenice od subjektivnog mišljenja.	Način izražavanja „jedna od retkih“ ukazuje da nije reč o činjenici, već o subjektivnom mišljenju studenta.
<i>Biće dobar profesor ako se toga bude oslobodila.</i>	Identifikovanje sentimenta	Potreban kontekst da bi se adekvatno izvršila anotacija.

Nakon inicijalne eksploracije probnog korpusa definisana je shema anotacije i anotaciona procedura, koje su detaljnije opisane u narednom poglavlju.

### 4.1.3. Kreiranje anotiranog korpusa za sentiment analizu pominjanja visokoškolskih institucija

Na osnovu prethodno modelovanog sadržaja izvedena je procedura anotacije, prikazana na slici 4.7, prema kojoj je realizovan postupak anotacije. Detaljniji opis postupka anotacije dat je u uputstvu za anotaciju (Prilog D). Uputstvo pokriva brojne primere anotacije pripremljene na osnovu rečenica koje su obuhvaćene u prethodnoj fazi istraživanja.



Slika 4.7. Anotaciona procedura

Cilj anotacione sheme, prikazane u Prilogu C, je da se obuhvate iskazani sentimenti usmereni ka aspektima koji su predmet recenzije studenta. Anotacionom shemom obuhvaćena je anotacija na nivou rečenice i na nivou sentiment izraza. Svakom segmentu anotacije anotator dodeljuje tačno jedan aspekt, sentiment polaritet i intenzitet iskazanog osećanja, izdvaja sentiment reči i označava negaciju kada za to postoje uslovi. Tabela 4.5. daje sumarni pregled atributa.

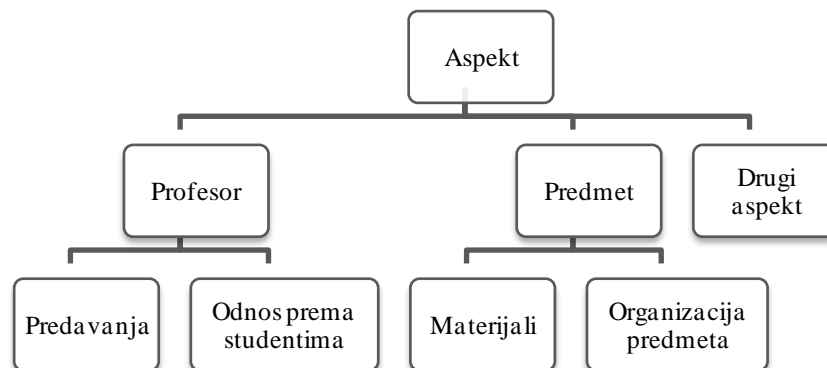
Tabela 4.5. Atributi definisani anotacionom shemom

Atribut	Upotreba	Vrednosti
Aspekt	Obavezno za svaku rečenicu. Definiše predmet o kom se piše.	Tačno jedna od predefinisanih vrednosti, slika 4.8.
Sentiment polaritet	Ukoliko je u rečenici iskazan lični stav ili mišljenje, atribut definiše polaritet.	Pozitivno, negativno, neutralno.
Intenzitet sentimenta	Definiše intenzitet sentiment iska za.	1, 2 ili 3.
Pozitivna reč	Opcioni atribut. Koristi se samo kada u rečenici u kojoj je iskazan pozitivan sentiment postoji pozitivna reč.	Domenski zavisna, ne postoji predefinisana lista.
Negativna reč	Opcioni atribut. Koristi se samo kada u rečenici u kojoj je iskazan negativan sentiment postoji negativna reč.	Domenski zavisna, ne postoji predefinisana lista.
Negacija	Opcioni atribut. Koristi se samo kada je u rečenici prisutna negacija.	Reči poput ne, nije, nema. Ne postoji predefinisana lista.
Opseg važenja negacije	Ukoliko je u rečenici prisutna negacija, atribut se koristi da bi označio početak i kraj delovanja negacije.	Tekst koji anotator odabira, zavisno od opsega delovanja negacije.

Aspekt i polaritet sentimenta mogu biti iskazani na implicitan i eksplicitan način. Iz tog razloga anotatori koriste svoja iskustvena saznanja i zdravorazumsko rasuđivanje kako bi utvrdili koji aspekt je predmet pisanja, odnosno koji polaritet sentimenta se pridružuje identifikovanom aspektu.

### Atribut Aspekt

Atribut *Aspekt* oslikava predmet komentarisanja. S obzirom na izvor podataka, prirodu sadržaja, kao i način na koji se iznose mišljenja, definisana su dva nivoa aspekta. Viši nivo se odnosi na *Profesora* i *Predmet*. Za oba aspekta na višem nivou definisane su potkategorije. Aspekt *Profesor* obuhvata dve kategorije: *Predavanja* i *Odnos profesora prema studentima*, dok aspekt *Predmet* obuhvata: *Materijale* i *Organizaciju predmeta*. Za rečenice čiji se sadržaj ne odnosi na domen visokog obrazovanja, ali ipak sadrži određeni sentiment, definisana je vrednost *Drugi aspekt*. Hijerarhija aspekata je predstavljena na slici 4.8.



Slika 4.8. Vrednosti atributa *Aspekt*

Proces anotiranja *Aspekta* zahteva da se komentaru prvo pridruži neki od aspekata na nižem hijerarhijskom nivou. Ukoliko ne postoji osnova za takav način anotacije, komentaru se dodeljuje aspekt višeg hijerarhijskog nivoa. Tabela 4.6. daje detaljan prikaz definisanih aspekata sa opisom i jednim ilustrativnim primerom. Detaljni primeri anotacije dati su u uputstvu za anotaciju (Prilog D.1). Ukoliko student u jednoj rečenici piše o različitim aspektima, svaki deo rečenice koji se odnosi na jedan aspekt anotira se posebno.

### Atribut Sentiment polaritet

Identifikovani sentiment koji student iskazuje u vidu uverenja, mišljenja, emocije, procene ili suda anotiran je postavljanjem vrednosti atributa *Sentiment polaritet*. Anotator označava rečenicu kao *pozitivnu*, *negativnu* ili *neutralanu*. Ukoliko student u rečenici koja je predmet anotacije iskazuje pozitivnu emociju, te se predmet o kojem piše može smatrati kao prednost ili

Tabela 4.6. Lista predefinisanih aspekata sa kratkim opisima i primerom

Aspekt	Opis	Primer
Profesor	Student piše o profesoru, ali ne opisuje predavanja koja profesor drži niti njegov odnos prema studentima. Student izostavlja subjekt u rečenici i navodi samo pohvalu ili pokudu.	Jeste malo kapriciozan, ali je ipak dobar profesor.
Predavanja	Student piše o načinu izlaganja profesora, načinu prenošenja znanja, koliko dobro profesor objašnjava gradivo koje predaje, stepen interesantnosti predavanja, upotrebljivosti predavanja za pripremu ispita, i slično.	Predavanja su retko kad živahna, ali su uvek korisna.
Odnos prema studentima	Student piše o ophođenju profesora prema studentima na nastavi, tokom provere znanja, na konsultacijama i u drugim situacijama, zatim o kvalitetu komunikacije između profesora i studenta, (ne)postojanju jasnog kriterijuma pri ocenjivanju, i slično.	Vraća radove jer u njenom imenu nije bilo dodato prezime Ilić, koje sad više nema.
Predmet	Student piše uopšteno o predmetu koji profesor predaje, ali ne opisuje nastavne materijale niti organizaciju predmeta.	Ceo predmet je bizaran, na kraju nemate nijedan pristojan rad, a nemate ni novo znanje.
Materijali	Student piše o dostupnosti i kvalitetu nastavnih materijala.	Materijal za učenje jeste dostupan, ali je nerazumljiv, nabacan, bez reda, bez nekog razumljivog rečnika.
Organizacija predmeta	Student piše o načinu na koji je nastava organizovana, logistici ispita, organizaciji polaganja predmeta (broj kolokvijuma) i sličnim tehničkim aspektima izvođenja nastave i provere znanja.	Svaki ispit je podeljen u glave, pri čemu se svaka glava polaže kao zaseban ispit, a ispitivanje traje i po više od 3, 4 sata...
Drugi aspekt	Predmet studentske recenzije ne pripada domenu visokog obrazovanja ili nema uticaja na (ne)kvalitet obrazovnog procesa.	Bitno da se ofarbala :).

poželjna osobina, rečenici se pridružuje pozitivan sentiment. Ukoliko student iskazuje negativnu emociju, te se predmet o kojem piše može smatrati nedostatkom ili nepoželjnom osobinom, rečenici se pridružuje negativan sentiment. U suprotnom, rečenica je neutralna. U pojedinim rečenicama je prisutan mešoviti sentiment. Takve rečenice se parcijalno anotiraju kako bi jedna anotacija oslikavala jedan sentiment. Uputstvo detaljno ilustruje primere anotacije različitih polariteta sentimenta ([Prilog D.2](#)).

### Atribut Intenzitet sentimenta

Identifikovani sentiment iskazuje različite intenzitete koji se anotiraju atributom *Intenzitet sentimenta*. Anotator definiše intenzitet sentimenta odabirom jedne od tri predefinisane vrednosti: 1 za slab intenzitet, 2 za srednji i 3 za jak intenzitet iskazanog sentimenta. Vrednost 3

će biti dodeljena ukoliko student govori u superlativu ili upotrebljeni izrazi značajno i jasno odstupaju od drugih iskaza sentimenta. Vrednost 2 se dodeljuje kada su istaknute pozitivne ili negativne osobine bez posebnog oduševljenja ili euforije, dok se slabi iskazi emocija označavaju vrednošću 1. U [Prilogu D.3.](#) dati su detaljni primeri anotacije rečenica u kojima su iskazana pozitivna i negativna osećanja različitog intenziteta.

### **Atribut Sentiment reč**

Jedna ili više reči (frazе) koje nose i oslikavaju identifikovani sentiment u rečenici označavaju se kao sentiment reči. Atributi *PozitivnaReč* i *NegativnaReč*, koji se dodeljuju sentiment rečima u zavisnosti od pozadiskog osećanja, objedinjeni su nazivom *Sentiment reč* mada predstavljaju dva odvojena atributa. U anotiranju sentiment reči potrebno je voditi se MIN strategijom (Vincze, et al., 2008), koja podrazumeva da se minimalna jedinica teksta izdvaja kao sentiment reč. U uputstvu za anotaciju korpusa dati su primeri anotacije sentiment reči i detaljna uputstva za njihov obuhvat ([Prilog D.4.](#)).

### **Anotacija negacije**

Anotacija negacije je svedena na označavanje ključnih reči negacije (ne, nije, nema i sl.) i opsega važenja negacije. Opseg važenja negacije se odnosi na set reči čije značenje negacija menja. Anotatori su upućeni na MIN/MAX strategiju predloženu u radu (Vincze, et al., 2008) koja podrazumeva da se minimalna jedinica teksta koja iskazuje negaciju označi kao ključna reč (dodeljivanjem atributa *Negacija*), dok se kao opseg važenja negacije obuhvata najduža moguća sintaksička jedinica. Opseg važenja negacije se obično prostire do interpunkcijskog znaka, veznika ili druge celine u rečenici. Detaljno uputstvo za anotaciju i primeri negacije dati su u [Prilogu D.5.](#)

#### **4.1.3.1. Proces anotacije korpusa**

Za realizaciju anotacije korpusa inicijalno je angažovano 6 anotatora, od kojih je jedan i autor rada. Odabranim anotatorima srpski jezik je maternji, visokoobrazovani su i poznaju proces studiranja, što ih je kvalifikovalo za realizaciju anotacije korpusa na srpskom jeziku za domen visokog obrazovanja.

Anotacija korpusa je vršena kroz više faza. U prvoj fazi anotacije anotatorima je predstavljeno uputstvo i date su detaljne smernice za rad, kao i uzorak korpusa (po 170 rečenica) nad kojim su se obučavali i nad kojim se testirao kvalitet uputstva i konzistentnost anotacije. U [Prilogu D.](#) date su detaljne informacije o uputstvu za anotaciju korpusa koje su četiri angažovana anotatora pratila, kao i detaljne informacije o anotiranim atributima. Tokom ove faze anotatori su podsticani na diskusiju i dodatna pitanja kako bi se razrešili problemi i nedoumice.

Identifikovani problemi, prikazani u tabeli 4.7, ukazali su na potrebu za dodatnim obučavanjem anotatora, kao i za određenim modifikacijama u anotacionoj shemi.

**Tabela 4.7. Identifikovani problemi u prvoj fazi anotacije**

<b>PROBLEM ASPEKTA</b>	
Profesor	Zamenjuje se aspektom Odnos prema studentima i obrnuto.
	Zamenjuje se aspektom Predavanja i obrnuto.
Organizacija predmeta	Neadekvatno identifikovanje.
	Zamenjuje se aspektom Predavanja.
Drugi aspekt	Neadekvatno identifikovanje.
	Neadekvatno anotiranje.
<b>PROBLEM NEUTRALNOG SENTIMENTA</b>	
Neadekvatno identifikovanje.	
Neadekvatno anotiranje.	
<b>PROBLEM SENTIMENT REČI</b>	
Nekompletno izdvajanje sentiment reči.	
Neoznačavanje reči.	
Prepričavanje sentiment reči.	
Neadekvatno označavanje.	
Nekompletno označavanje reči.	
<b>OSTALI PROBLEMI</b>	
Obuhvat teksta	Izostavljanje početnog slova.
	Neobuhvatanje interpunkcijskih znakova.
	Označavanje istog karaktera u različitim anotacijama.
Nekonzistentost u anotacijama	

Većina problema je bila jedinstvena za pojedinačnog anotatora. Kako se može primetiti iz tabele, anotatori nisu u potpunosti razumeli definicije pojedinih aspekata usled čega je došlo do njihovog neadekvatnog identifikovanja ili zamene drugačijim aspektom. Primeri P.1. i P.2. ilustruju zamenu aspekta *Profesor* i *Odnos prema studentima*, i obrnuto. Primeri P.3. i P.4. ilustruju zamenu aspekta *Profesor* i *Predavanja*, i obrnuto. Primeri P.5-P.7. ilustruju zamenu aspekta *Organizacija predmeta* i *Predavanja*. Ovo je jasno ukazalo na potrebu za dodatnim obučavanjem anotatora, kao i za modifikovanjem uputstva kako bi se jasnije opisali definisani aspekti i postavila jasna granica između aspekta *Profesor* i *Odnos prema studentima*, *Profesor* i *Predavanja*, kao i *Predavanja* i *Organizacija predmeta*.

(P.1.) *Mnogo fin, kulturnan, uvek izađe u susret.*

(P.2.) *Sa ovim profesorom je teško komunicirati.*

(P.3.) *Ponekad priča neumesne stvari.*

(P.4.) *Zanimljivo predaje, sve što pripremi da se na predavanju uradi, to se i ostvari.*

(P.5.) *Na časovima čita tekstove, po nekoliko rečenica, i onda ih analizira na način na koji niste mogli naći ni u jednom udžbeniku.*

(P.6.) *Vrlo interesantna predavanja, praktične vežbe koje gađaju na inventivnost i drugačiji pogled na svet...*

(P.7.) *Vežbe su za svaku pohvalu – i korisne i praktične.*

*Drugi aspekt* je predstavljao posebno kritičnu kategoriju. Ukoliko je u rečenici izostao subjekat i studenti napišu samo pohvalu ili pokudu (primer: *Sjajan!* ili *Loš.*) anotatori su učestalo anotirali ovakve rečenice kao *Drugi aspekt*. S obzirom na prirodu sajta i činjenicu da se komentari ostavljaju prvenstveno kako bi se ocenio profesor, u većini ovih primera subjekat je podrazumevan i odnosi se na profesora. Dodatno, kada bi identifikovali *Drugi aspekt* u rečenici, anotatori su učestalo izostavljali ostale attribute. Kako je u fokusu istraživanja iskazani sentiment u komentarima, rečenicama koje se odnose na drugi aspekt takođe je bilo potrebno dodeliti sentiment polaritet, intenzitet i izdvojiti sentiment reč. Identifikovanje *Drugog aspekta* je takođe predstavljalo problem. Primer (P.8.) ilustruje rečenicu koja je pogrešno anotirana kao drugi aspekt. Uputstvo za anotaciju je prošireno pojašnjenjem *Drugog aspekta*, a anotatori su dodatno obučeni.

(P.8.) *Znam da je većina mojih kolega imala problem oko seminarskih radova.*

Anotatori su iskusili problem da adekvatno prepoznaju neutralni sentiment. Primeri (P.9 - P.11) ilustruju pogrešno označen neutralni sentiment. Takođe, u pojedinim rečenicama nakon identifikovanja neutralnog sentimenta (ispravno ili neispravno) anotatori su dodeljivali vrednosti za intenzitet iskazanog sentimenta. Neutralnim sentimentom se označavaju rečenice u kojima nema ni pozitivnog ni negativnog sentimenta, dakle, ne postoji osećanje, samim tim, nema ni intenziteta iskazanog osećanja. Anotatori su najčešće pogrešno anotiranim neutralnim rečenicama dodali i oznaku intenziteta iskazanog sentimenta, što je i predstavljao jasan vodič u budućem radu: rečenica kojoj imaju potrebu da dodele intenzitet iskazanog sentimenta ne može biti neutralna. O problematičnim primerima i načinu anotacije diskutovano je sa anotatorima.

(P.9.) *Za sada je korektan.*

(P.10.) *Ponekad se previše udubi u objašnjavanje primera, pa ne stigne skoro ni do pola neke lekcije.*

(P.11.) *Literatura ne postoji.*

Inicijalnom anotacionom shemom bio je definisan jedan atribut *SentimentReč* kojem su anotatori dodeljivali vrednost upisivanjem konkretne reči u predefinisano polje, bez obzira na polaritet sentimenta. Ovo je predstavljao veliki propust jer je anotatorima davana sloboda da sentiment reči pišu u slobodnom stilu (u primeru P.12. kao sentiment reč je izdvojeno *čitanje skripte*, dok



je u primeru P.13. izdvojeno *nerazumno predavanje*) i onemogućeno je da izdvoje više od jedne sentiment reči u rečenici.

(P.12.) *Profesorka čita word dokument svoje skripte (koja je ujedno loše napisana) na predavanju.*

(P.13.) *Gradivo je nerazumljivo ili to profesor tako prezentuje.*

Kao problem se učestalo pojavljivalo neizdvajanje sentiment reči u rečenicama u kojima se one jasno mogu prepoznati. Anotatori su označavali lična imena kao sentiment reči ili su u sentiment reč uključivali povratnu zamenicu, što je zahtevalo dodatno pojašnjenje i obuku. Označavan je duži deo teksta nego što je potrebno, na primer *odokativno ocenjivanje* umesto *odokativno, lepo objašnjava* umesto *lepo*. U rečenicama u kojima je prisutna negacija anotatori su učestalo označavali sentiment reči nedovoljno obuhvatno, na primer *korektna* umesto *nije korektna* ili *ocenjuje* umesto *ne ocenjuje*, o čemu se dodatno diskutovalo sa anotatorima. Ovi problemi su doveli do promene u pristupu anotaciji, pre svega, uvedena su dva odvojena obeležja kojima će se posebno označavati pozitivna, odnosno negativna sentiment reč. Vrednosti ovih atributa se više ne postavljaju upisivanjem, nego selektovanjem konkretne sentiment reči u tekstu, kojoj se potom dodeljuje vrednost *PozitivnaReč* ili *NegativnaReč*. Na ovaj način je prevaziđen problem prepričavanja sentiment reči i omogućeno je odvajanje više reči u jednoj rečenici. Takođe, uvedena je MIN strategija u označavanju sentiment reči, koja podrazumeva da se minimalna jedinica teksta označi kao sentiment reč. U ovoj fazi je uvedena anotacija negacije, koja je zahtevala dodatno obučavanje i upoznavanje anotatora sa novim atributima i oznakama.

Takođe, radilo se i na usklađivanju i standardizovanju anotacije, te su anotatori upućeni na uravnoteženo označavanje teksta koji je predmet anotacije (da ne izostavljaju slova i interpunkcijske znake) i da povedu računa o konzistentnosti anotacija (da istoj rečenici ne dodeljuju različite vrednosti, P.14).

(P.14.) *Uvek spremna da pomogne. – intenzitet: 2*

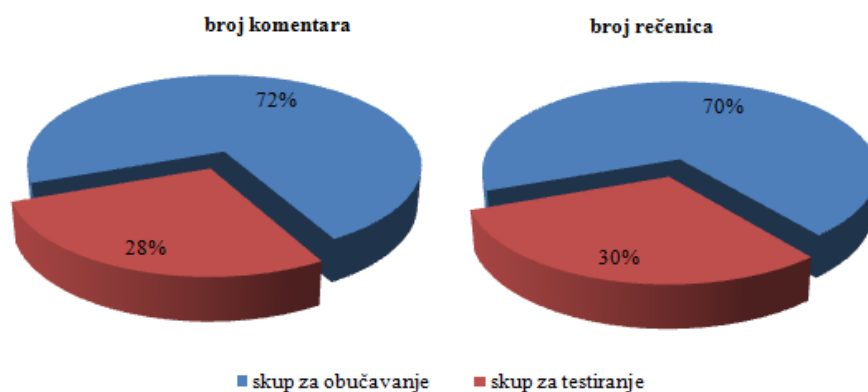
*Spreman uvek da pomogne. – intenzitet: 3*

Kako je ova faza bila vremenski zahtevna, a anotatori su morali da ulože značajan napor da bi savladali anotacionu proceduru, dva anotatora su odustala, te su anotiranje celokupnog korpusa realizovala 4 anotatora<sup>39</sup>, kojima je u narednoj fazi prezentovana nova anotaciona shema i realizovano je dodatno obučavanje i usklađivanje. U drugoj fazi anotatorima je prezentovano 50 najproblematičnijih rečenica iz prethodne faze, koje su ujedno obuhvatale i primere rečenica sa negacijom. Sa anotatorima se radilo i na načinu prepoznavanja višestrukih aspekata i sentimenta

<sup>39</sup> Alenka Grljević Svilengaćin, diplomirani pravnik. Gordana Svilengaćin, master profesor jezika i književnosti (srbista). Gordana Somborski, diplomirani ekonomista. Olivera Grljević, diplomirani ekonomista – master, angažovana u kreiranju zlatnog standarda.

u rečenici, kao na njihovom pravilnom anotiranju. Neadekvatno razdvajanje ovakvih rečenica, nastalo prevashodno iz lošeg poimanja pojedinih aspekata, predstavljalo je još jedan od učestalih problema u prvoj fazi anotacije. Problemi koji su se javili u prethodnoj fazi anotacije prevaziđeni su u drugoj fazi. Međutim, u drugoj fazi anotiranja pojavila se nekonzistentnost u preraspodeli rečenica za anotaciju. Pojedini anotatori bi u jednoj rečenici prepoznali 2 ili 3 aspekta/sentimenta, dok bi drugi u istoj prepoznali manje. Sa anotatorima je detaljno diskutovano o ovim primerima i izvršena je njihova analiza, kao i dodatno pojašnjavanje aspekta *Organizacija predmeta* i *Odnos prema studentima*.

Kako je većina problema otklonjena i anotatori su postali usklađeniji u svojim anotacijama, pripremljeni korpus je podeljen na skup za obučavanje i skup za testiranje, slika 4.9. Skup za obučavanje se koristi za obučavanje algoritama, dok se skup za testiranje koristi za analizu preciznosti rada algoritama nad nepoznatim podacima, te je u ovoj fazi neophodno izvršiti raspodelu korpusa kako bi se deo sadržaja izolovao iz faze obučavanja i ostao nepoznat algoritmima. U [Poglavlju 4.1.6.](#) prikazani su rezultati rada razvijenih modela nad skupom za obučavanje, kao i preciznost rada nad nepoznatim podacima – skupom za testiranje. Rečenice koje su obuhvaćene prethodnim fazama istraživanja i o kojima se detaljno diskutovalo sa anotatorima predstavljaju deo skupa za obučavanje. Anotator koji je bio angažovan u kreiranju uputstva za anotaciju izvršio je konačnu presudu o anotacijama ovih rečenica pre njihovog uključivanja u konačni korpus. Takođe, iz skupa za obučavanje je nasumično izdvojeno 15.36% komentara sa kojima do ovog momenta nije rađeno (430 od ukupno 2800 komentara koji čine skup za obučavanje) i koji su u narednoj fazi prezentovani anotatorima kako bi se nakon anotacije izvršila ocena saglasnosti njihovih anotacija. U ovoj fazi anotatori su radili anotaciju sadržaja potpuno samostalno i bez konsultacija. Nakon procene stepena saglasnosti anatora (rezultati su detaljnije prikazani u narednom poglavlju) usledile su još dve faze anotacije: prva – kojom se dovršilo anotiranje skupa za obučavanje, i druga, poslednja faza, – u kojoj se anotirao skup za testiranje.



Slika 4.9. Ras podela korpusa na skup za obučavanje i testiranje

#### 4.1.3.2. Procena stepena saglasnosti anotatora

Anotaciona shema se koristi kako bi se označile i pridružile dodatne informacije podacima, a prevashodno informacije o emocijama. Kako su emocije vrlo subjektivna kategorija, potrebna je objektivna mera kao dokaz primenljivosti definisane sheme koju drugi istraživači mogu da razumeju i na osnovu koje mogu pouzdano suditi o istraživanju, kao i korektnosti modelovanog fenomena (Siegert, et al., 2014). Za ove potrebe podesnu meru predstavlja stepen saglasnosti anotatora. Ona određuje stepen u kojem su dva ili više anotatora postigla iste rezultate u anotiranju određenog teksta. Da bi se ručno anotirani podaci koristili, potrebno je pokazati da su takvi podaci **pouzdati** (Artstein, et al., 2008). Kako autori ističu, podaci se smatraju pouzdanim ukoliko se među anotatorima dokaže postojanje **slaganja** po pitanju kategorija dodeljenih jedinicama teksta. Ukoliko različiti anotatori dobijaju dosledno slične rezultate, može se zaključiti da su slično shvatili smernice za anotaciju i očekivati da će konzistentno obavljati zadatak anotacije. Pouzdanost tako predstavlja preduslov za demonstriranje **validnosti** anotacione sheme, te ona oslikava istinu fenomena koji je predmet istraživanja. Neslaganje između anotatora može ukazati na pogrešno razumevanje uputstva za anotaciju, kao i na neadekvatnost anotacione sheme. Postizanje slaganja između anotatora i dalje ne osigurava validnost. Anotatori mogu deliti isto mišljenje, a da u isto vreme nisu u pravu. Za kalkulisanje saglasnosti anotatora najčešće se koriste kappa statistike: ukoliko je angažovano dva anotatora *Cohen kappa*  $\kappa$  (Cohen, 1960), za više od dva anotatora *Fleiss kappa*  $K$  (Fleiss, 1971), (Fleiss, et al., 2003)) i *Krippendorff alfa* (Krippendorff, 2012), kojom se ocenjuje stepen neslaganja anotatora.

Ukoliko anotaciju radi više anotatora, njihova saglasnost se ne može posmatrati kao procenat anotacija u kojima je prisutna saglasnost, jer će postojati anotacije u kojima će jedan ili više anotatora odstupiti. U radu (Fleiss, 1971) predloženo je merenje stepena saglasnosti anotacije stavke kao udela saglasnih anotacija u ukupnom broju anotacija za konkretnu stavku, kako je objašnjeno u nastavku.

Ako se  $N$  definiše kao ukupan broj anotacija,  $n$  kao broj anotatora,  $k$  kao ukupan broj kategorija (vrednosti atributa) koje se dodeljuju rečenicama,  $i$ , gde je  $i=1, \dots, N$  predstavlja pojedinačnu anotaciju (predmet anotacije) i  $j$ , gde je  $j=1, \dots, k$  predstavlja pojedinačnu kategoriju,  $n_{ij}$  predstavlja broj anotatora koji su  $i$ -toj anotaciji dodeliti  $j$ -tu kategoriju,  $p_j$ , koje predstavlja proporciju svih anotacija za  $j$ -tu kategoriju, računa se prema sledećoj formuli:

$$p_j = \frac{1}{Nn} \sum_{i=1}^N n_{ij}$$

U ovoj fazi istraživanja anotatori  $n=4$  dodeljuju jednu od  $k=3$  kategorije za sentiment polaritet ( $k=7$  za atribut aspekt, odnosno  $k=3$  za atribut intenzitet sentimenta) rečenicama  $N=568$  (izuzev

u ocenjivanju sentiment intenziteta, gde je  $N$  umanjen za broj neutralnih rečenica kojima se ne dodeljuje intenzitet sentimenta i iznosi  $N=563$ ). U nastavku je ilustrovana kalkulacija stepena saglasnosti anotatora za anotacije atributa Sentiment polaritet. Detaljniji prikaz kalkulacija je dat u [Prilogu E](#). Sveukupne proporcije za vrednosti atributa sentiment polaritet su  $p_{(\text{pozitivan})} = 0.616$ ,  $p_{(\text{negativan})} = 0.382$  i  $p_{(\text{neutralan})} = 0.002$ . Kako je  $\sum_{i=1}^N n_{ij} = n$ ,  $\sum_j p_j = 1$ .

Stepen saglasnosti između  $n$  anotatora za  $i$ -tu kategoriju definiše se kao proporcija saglasnih parova u svim mogućim  $n(n-1)$  parovima anotacije:

$$P_i = \frac{1}{n(n-1)} \sum_{j=1}^k n_{ij}(n_{ij} - 1) = \frac{1}{n(n-1)} \left( \sum_{j=1}^k n_{ij}^2 - n \right)$$

Tabela 4.8. ilustruje rezultujuće vrednosti za prvih 6 anotacija u kojima se mogu zapaziti oscilacije u stepenima saglasnosti anotatora.

**Tabela 4.8. Stepen saglasnosti anotatora za i-tu kategoriju**

i	Pozitivan (j=1)	Negativan (j=2)	Neutralan (j=3)	$P_i$
1	1	3	0	0.50
2	1	3	0	0.50
3	1	2	1	0.17
4	1	3	0	0.50
5	0	3	1	0.50
6	0	4	0	1.00

Sveukupni stepen saglasnosti se meri kao srednja vrednost svih saglasnosti po kategorijama,  $P_i$ :

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i = \frac{1}{Nn(n-1)} \left( \sum_{i=1}^N \sum_{j=1}^k n_{ij}^2 - Nn \right)$$

Za atribut Sentiment polaritet  $\bar{P} = 0,9830$ , a može se interpretirati na sledeći način. Neka je nasumično odabrana rečenica anotirana od strane nasumično odabranog anotatora. Ukoliko se rečenica takođe anotira od strane drugog nasumično odabranog anotatora, druga anotacija bi bila u saglasnosti sa prvom u preko 98% slučajeva.

Određeni stepen saglasnosti se može očekivati i na osnovu slučajnosti. Ukoliko anotatori čine anotacije slučajno, srednja vrednost saglasnosti bi bila:

$$\bar{P}_e = \sum_{j=1}^k p_j^2$$

Rezultujuća vrednost za atribut Sentiment polaritet je  $\bar{P}_e = 0,5253$ .

U konačnoj formuli *Fleiss kappa*, koja je prikazana u nastavku,  $1 - \bar{P}_e$  prikazuje za koliko je izračunata saglasnost veća ili manja od očekivane (nasumične) saglasnosti. Normalizovana mera sveukupne saglasnosti, korigovana za stepen očekivane saglasnosti, izražava se kao količnik razlike  $\bar{P} - \bar{P}_e$  u odnosu na  $1 - \bar{P}_e$ .

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

Paralelni prikaz rezultujućih vrednosti za sve anotirane attribute prikazan je u tabeli 4.9, nakon koje sledi detaljno objašnjenje dobijenih vrednosti. U radu (Fleiss, 1971) dalje je definisano da se odsustvo stvarne saglasnosti između  $n$  anotatora tumači kao njihova nemogućnost razlikovanja jednog predmeta anotacije od drugog, što implicira da anotatori nad svakim predmetom anotacije primenjuju sveukupne stope saglasnosti ( $p_j$ ). Pod pretpostavkom nepostojanja saglasnosti, ovo dalje znači da  $n$  anotacija nad svakim predmetom anotacije predstavljaju varijable sa verovatnoćama  $p_1, \dots, p_k$ . Kada je  $N$  dovoljno velik, suma kvadriranih vrednosti anotacija za  $i$ -ti predmet anotacije će biti nezavisna od sume kvadriranih vrednosti za  $i'$ -ti predmet anotacije. U uslovima kada je  $N$  velik, varijansa  $\kappa$ , pod pretpostavkom nepostojanja saglasnosti osim nasumične, definisana je na sledeći način:

$$\text{Var}_{\kappa} = \frac{2}{Nn(n-1)} \times \frac{\sum_j p_j^2 - (2n-3)(\sum_j p_j^2)^2 + 2(n-2)\sum_j p_j^3}{(1 - \sum_j p_j^2)^2}$$

Kako bi se testirala pretpostavka nezavisnosti anotatora (kappa je nula), standardna greška kappa-e (SE) se dobija kao:

$$SE_{(\kappa)} = \sqrt{\text{Var}_{\kappa}}$$

Hipoteza da je saglasnost među anotatorima veća od slučajne, izračunate kao:

$$z = \frac{\kappa}{SE_{\kappa}}$$

potvrđuje se poređenjem dobijene vrednosti sa tabelarnim vrednostima standardne normalne distribucije, a odbacuje se u slučaju da je  $z$  dovoljno veliko. Jednostrani test je u ovom slučaju pogodniji od dvostranog.

Rezultujuće vrednosti za sve anotirane attribute prikazane su u tabeli 4.9.

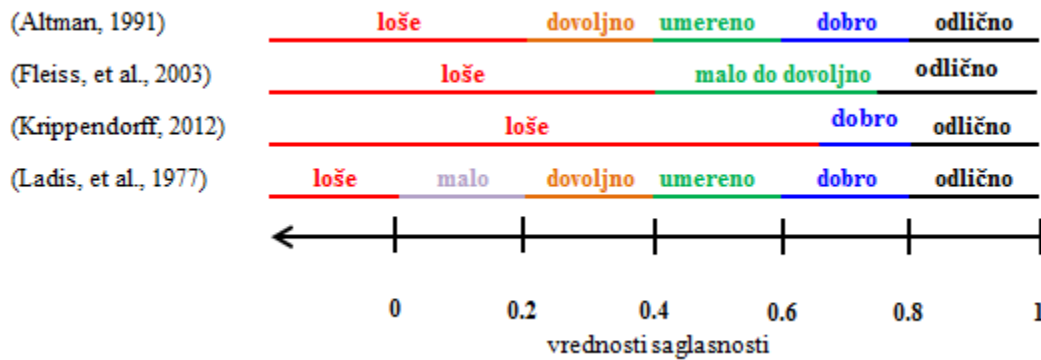
**Tabela 4.9. Saglasnost anotatora po atributima**

	$P$	$P_e$	$\kappa$	$SE(\kappa)$	$z$
<b>Sentiment polaritet</b>	0.983	0.525	<b>0.964</b>	0.035	27.189
<b>Aspekt</b>	0.915	0.591	<b>0.792</b>	0.052	15.350
<b>Sentiment intenzitet</b>	0.749	0.549	<b>0.443</b>	0.042	10.571

Kappa može poprimiti vrednosti od -1 do 1, gde -1.0 označava potpuno neslaganje koje nije rezultat slučajnosti, 0.0 označava slaganje jednako očekivanom slučajnom slaganju, a 1.0 je indikacija potpunog slaganja koje nije rezultat slučajnosti. Postoji više pristupa interpretiranju rezultujuće *kappa* statistike, koji su prikazani na slici 4.10. (Siegert, et al., 2014). U praksi se najčešće primenjuju smernice za interpretaciju koje su dali (Ladis, et al., 1977), prema kojima se može zaključiti da je postignuta odlična saglasnost anotatora za atribut Sentiment polaritet ( $\kappa = 0.964$ ), dobra saglasnost za atribut Aspekt ( $\kappa = 0.792$ ) i umerena saglasnost za atribut Intenzitet sentimenta ( $\kappa = 0.443$ ), koja ukazuje na poteškoće u anotiranju ovog atributa od strane ljudi. Rezultati saglasnosti anotacija atributa Sentiment polaritet i Aspekt zadovoljavajući su, dok su prihvatljivi za Intenzitet sentimenta, čime je potvrđen kvalitet uputstva za anotaciju. U poređenju sa drugim istraživanjima dobijena je nešto viša saglasnost u anotiranju polariteta sentimenta:

- Saglasnost u označavanju polariteta na nivou rečenice je  $\kappa = 0.92$ , na nivou sentiment izraza  $\kappa = 0.94$  (Toprak, et al., 2010).
- Saglasnost anotatora u označavanju polariteta sentimenta na nivou rečenice iznosi 70-79% (Gamon, et al., 2005).
- Za pozitivan sentiment postignuta saglasnost anotatora – kappa – je iznad 0.8, za negativan 0.78, dok za neutralan i konfliktni sentiment iznad 0.6 (Ganu, et al., 2009).
- Saglasnost anotatora u označavanju pozitivnog sentimenta kreće se između 0.6-0.84, dok je izuzetno niska u označavanju negativnog sentimenta 0.4-0.45 (Zirn, et al., 2011).

Ovako visoka saglasnost u anotiranju polariteta sentimenta može se pripisati: a) dužini tekstova koji su bili predmet anotacije i b) sentiment rečima. Naime, tekstovi su kratki, neke rečenice imaju samo jednu reč, što značajno olakšava prepoznavanje sveukupnog sentimenta. U ovako kratkom tekstu prisutne se i određene fraze ili kombinacije reči koje se pojavljuju dovoljno često da se mogu svrstati u fraze, poput: bila mi je čast, svaka čast, sve pohvale, jedan od najboljih, i slično, a koje jasno govore o iskazanom sentimentu.



Slika 4.10. Uporedni prikaz interpretacije kapa statistike (Siegert, et al., 2014)

Prema našem saznanju komparativno istraživanje sprovedeno je u radu (Ganu, et al., 2009), u kojem su eksplicitno anotirani aspekti, odnosno predmeti pisanja u recenzijama restorana. Prema vrednostima kapa statistike opisanim u ovom istraživanju postignuta saglasnost u anotiranju atributa Aspekt je približna, naime u radu (Ganu, et al., 2009) za aspekte *Hrana*, *Cena* i *Usluga* postignuta je saglasnost iznad 0.8, za *Razno* i *Ambijent* iznad 0.6, dok je za aspekt *Anegdote* postignuto umereno slaganje 0.51, koje su identifikovali kao problematičnu kategoriju. Detaljniji prikaz saglasnosti anotacija po pojedinim kategorijama atributa Sentiment polaritet, Aspekt i Intenzitet sentimenta prikazuje tabela 4.10.

Procena intenziteta iskazanog sentimenta podložna je subjektivnom utisku i nije moguće u potpunosti je standardizovati instrukcijama. Iz ovog razloga najniži stepen saglasnosti postignut je upravo za ovaj atribut. Međutim, i druga istraživanja ukazuju na sličan problem. Npr. u radu (Toprak, et al., 2010) postignuta je saglasnot od 0.55.

Ako nasumično odabran anotator svakom predmetu anotacije dodeljuje jednu od kategorija  $k$  i ponovo se anotiraju od strane drugog nasumično odabranog anotatora samo oni predmeti anotacije kojima je dodeljena  $j$ -ta kategorija, bazični indeks saglasnosti predstavlja uslovnu verovatnoću da je druga anotacija karegorija  $j$  s obzirom na to da je prva anotacija bila  $j$ -ta kategorija (Fleiss, 1971). Otuda se stepen saglasnosti anotatora u dodeljivanju kategorije  $j$  predmetu anotacije (rečenici) računa na sledeći način:

$$P_j = \frac{\sum_{i=1}^N n_{ij}(n_{ij} - 1)}{\sum_{i=1}^N n_{ij}(n - 1)} = \frac{\sum_{i=1}^N n_{ij}^2 - Nnp_j}{Nn(n - 1)p_j}$$

Pod pretpostavkom da stepen slaganja nije manji od nasumičnog, tj. slučajnog, očekivalo bi se da je ova verovatnoća jednaka verovatnoći (ne uslovnoj) pridruživanja predmeta anotacije kategoriji  $j$ , u oznaci  $p_j$ . Mera stepena slaganja prilikom pridruživanja kategoriji  $j$ , koja je manja od slučajnog slaganja u tom slučaju se računa kao:

$$\kappa_j = \frac{\bar{p}_j - p_j}{1 - p_j} = \frac{\sum_{i=1}^N n_{ij}^2 - Nnp_j[1 + (n-1)p_j]}{Nn(n-1)p_jq_j}$$

gde je  $q_j = 1 - p_j$ . Mera sveukupnog slaganja, u oznaci  $\kappa$ , predstavlja ponderisani prosek  $k_j$  za svaku kategoriju:

$$\kappa = \frac{\sum_j p_j q_j k_j}{\sum_j p_j q_j}$$

Varijansa  $k_j$  se izračunava kao:

$$\text{Var}_{(\kappa_j)} = \frac{[(1 + 2(n-1)p_j)^2 + 2(n-1)p_jq_j]}{Nn(n-1)^2p_jq_j}$$

Hipoteza se može testirati upoređivanjem dobijene vrednosti

$$\frac{\kappa_j}{SE_{(\kappa_j)}} = \frac{\kappa_j}{\sqrt{\text{Var}_{(\kappa_j)}}}$$

sa tabličnim vrednostima standardne nominalne distribucije.

Tabela 4.10. daje sumarni pregled rezultujućih vrednosti po mogućim kategorijama svakog atributa. Kako neutralni sentiment nije bio zastupljen u uzorku tekstova koji su se u ovoj fazi istraživanja koristili, saglasnost anotatora u anotiranju ovog polariteta iznosi  $\kappa = 0$ . Sa druge strane, može se primetiti da je uravnotežena i visoka saglasnost anotatora u anotiranju pozitivnog i negativnog sentimenta. U fazi obučavanja anotatori su iskusili najviše poteškoća u anotiranju rečenica koje se odnose na *Organizaciju predmeta* i *Drugi aspekt*. Iz tabele 4.10. se vidi da se upravo u ovim aspektima javlja niža, odnosno umerena saglasnost – prema smernicama za interpretiranje kappa statistike, koje su date u radu (Ladis, et al., 1977). Sveukupna kappa za Intenzitet sentimenta je 0,443, međutim, na nivou pojedinih vrednosti ovog atributa anotatori postižu višu saglasnost ( $\kappa = 0,528$ ) kada su u pitanju rečenice u kojima je iskazan snažan sentiment, dok je niža saglasnost postignuta u anotiranju manje intenzivnog sentimenta.



Tabela 4.10. Saglasnost anotatora po vrednostima atributa

	$\bar{P}_j$	$\kappa_j$	Var( $\kappa_j$ )	SE( $\kappa_j$ )	Z
<b>SENTIMENT POLARITET</b>					
<b>Pozitivan</b>	0.990	0.973	0.005	0.070	13.836
<b>Negativan</b>	0.978	0.965	0.003	0.050	19.154
<b>Neutralan</b>	0	0	0.023	0.152	0
<b>ASPEKT</b>					
<b>Profesor</b>	0.949	0.796	0.008	0.091	8.755
<b>Predavanja</b>	0.814	0.795	0.002	0.041	19.291
<b>Odnos prema studentima</b>	0.827	0.802	0.002	0.041	19.666
<b>Predmet</b>	0.730	0.728	0.006	0.079	9.210
<b>Materijali</b>	0.952	0.952	0.006	0.079	12.048
<b>Organizacija predmeta</b>	0.519	0.517	0.013	0.115	4.482
<b>Drugi aspekt</b>	0.462	0.458	0.009	0.097	4.706
<b>INTENZITET SENTIMENTA</b>					
<b>Jak</b>	0.856	0.528	0.006	0.081	6.554
<b>Srednji</b>	0.536	0.378	0.002	0.044	8.559
<b>Slab</b>	0.362	0.326	0.002	0.045	7.273

Anotacija atributa PozitivnaReč, NegativnaReč i negacije rađena je na nivou izraza. Anotatori prema sopstvenom nađenju mogu izdvojiti jednu ili više sentiment reči u jednoj rečenici, što rezultuje drugačijim brojem anotacija, kao i drugačije izdvojenim rečima. Iz ovog razloga *Fleiss kappa* se ne može primeniti. Kako je istaknuto u radovima ((Wiebe, et al., 2005), (Kessler, et al., 2009) i (Pontiki, et al., 2014)) u uslovima kada anotatori mogu označiti različite izraze, adekvatnije je upotrebiti *agr* metriku, koja meri preklapanje anotacija (tzv. *recall*) u oznaci *R* kao meru slaganja anotacija dva anotatora, označenih sa *a* i *b*, pri čemu se njihovi skupovi anotiranih izraza označavaju sa *A* i *B* respektivno. Dakle, ukoliko *A* i *B* predstavljaju skupove anotiranih izraza od strane *a* i *b* anotatora, respektivno, a *A* predstavlja zlatni standard, metrika *agr* meri preklapanje definisano na sledeći način:

$$agr(a||b) = R = \frac{|A \cap B|}{|A|}$$

Kada se skup *B* definiše kao zlatni standard, metrika *agr* iskazuje preciznost anotacije *P* (engl. *precision*), kada se upoređuje broj anotacija koje je anotator *b* anotirao identično kao anotator *a* spram ukupnog skupa njegovih anotacija *B*.

$$agr(b||a) = P = \frac{|B \cap A|}{|B|}$$

Da bi se dobila jedna metrika performanse, računa se harmonijska sredina preciznosti i preklapanja iskazana F-merom:

$$F - mera = \frac{2 \times P \times R}{(P + R)}$$

Kako ilustruje tabela 4.11, postignuta je mnogo viša saglasnost u anotiranju pozitivnih sentiment iskaza (85.15% – 90.58%) u odnosu na negativne (68.01% - 84.91%). Kako ne postoji dvosmislenost u njihovom značenju, saglasnost u anotiranju ključnih reči negacije izuzetno je visoka, dok je kod delokruga njenog važenja postignuta nešto niža saglasnost kao posledica obuhvata kraćeg ili dužeg dela teksta. Sveukupno posmatrano, postignuta je dobra saglasnost u anotiranju negativne reči i odlična za ostale anotacije. Primera radi, anotator A2 je saglasan sa 88.57% označenih pozitivnih reči od strane anotatora A1, dok je anotator A1 saglasan sa 92.70% anotacija pozitivnih reči koje je uradio anotator A2.

Tabela 4.11. Saglasnost anotatora u anotiranju na nivou izraza

	Pozitivna reč			Negativna reč			Negacija			Opseg važenja negacije		
	P	R	F-mera	P	R	F-mera	P	R	F-mera	P	R	F-mera
A <sub>1</sub> - A <sub>2</sub>	0.9270	0.8857	0.9058	0.7547	0.7714	0.7630	0.9877	0.8989	0.9412	0.9012	0.8295	0.8639
A <sub>1</sub> - A <sub>3</sub>	0.9625	0.7965	0.8717	0.9490	0.7683	0.8491	1.0000	0.8539	0.9212	0.9342	0.8068	0.8659
A <sub>1</sub> - A <sub>4</sub>	0.9277	0.8450	0.8844	0.8309	0.7175	0.7700	0.9875	0.8876	0.9349	0.9000	0.8182	0.8571
A <sub>2</sub> - A <sub>3</sub>	0.9368	0.8114	0.8696	0.8745	0.6925	0.7730	0.9342	0.8765	0.9045	0.8289	0.7778	0.8025
A <sub>2</sub> - A <sub>4</sub>	0.8723	0.8316	0.8515	0.7426	0.6273	0.6801	0.9125	0.9012	0.9068	0.8375	0.8272	0.8323
A <sub>3</sub> - A <sub>4</sub>	0.8191	0.9016	0.8584	0.7463	0.7961	0.7704	0.9125	0.9605	0.9359	0.8500	0.8947	0.8718

Po završetku anotacije sveukupnog korpusa prema definisanim instrukcijama potrebno je doneti odluku o paralelnim anotacijama u smislu konačne presude o anotaciji koja će biti uvrštena u korpus. Na ovaj način se formira korpus za obučavanje i testiranje algoritama mašinskog učenja, takozvani *zlatni standard*. Kako je sugerisano u radu (Pustejovsky, et al., 2012), anotator koji je bio angažovan u kreiranju uputstva za anotaciju vršio je presudu o konačnoj anotaciji rečenica koje su u ovoj fazi istraživanja anotirane. Presuda se vršila na osnovu anotacije većine anotatora. Slika 4.11. ilustruje rezultat anotacije jednog komentara.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <XML>
- <TEXT>
  <![CDATA[ ne dolazi na predavanja, ne objasni kako i šta treba i posle se hvali kako niko
    ništa ne zna ...
  ]]>
</TEXT>
- <TAGS>
  <Aspekt id="A0" start="1" end="26" text="ne dolazi na predavanja," OznakaKomentara="317"
    vrednosti="profesor" SentimentPolaritet="negativan" Intenzitet="3" komentar="" />
  <Aspekt id="A1" start="27" end="54" text="ne objasni kako i šta treba" OznakaKomentara="317"
    vrednosti="odnos prema studentima" SentimentPolaritet="negativan" Intenzitet="3" komentar="" />
  <Aspekt id="A2" start="55" end="98" text="i posle se hvali kako niko ništa ne zna ..."
    OznakaKomentara="317" vrednosti="profesor" SentimentPolaritet="negativan" Intenzitet="3"
    komentar="" />
  <NegativnaRec id="N0" start="1" end="11" text="ne dolazi" />
  <NegativnaRec id="N1" start="27" end="37" text="ne objasni" />
  <NegativnaRec id="N2" start="66" end="71" text="hvali" />
  <Negacija id="Ne0" start="1" end="4" text="ne" />
  <Negacija id="Ne1" start="27" end="29" text="ne" />
  <Negacija id="Ne2" start="88" end="90" text="ne" />
  <OpsegVazenjaNegacije id="O0" start="5" end="25" text="dolazi na predavanja" />
  <OpsegVazenjaNegacije id="O1" start="30" end="54" text="objasni kako i šta treba" />
  <OpsegVazenjaNegacije id="O2" start="91" end="94" text="zna" />
</TAGS>
</XML>

```

Slika 4.11. Prikaz anotiranog komentara

Naredno poglavlje daje detaljan prikaz rezultujućeg korpusa, distribuciju vrednosti pojedinih atributa u korpusu, kao i zapažanja do kojih se došlo tokom anotacije o stilu pisanja, načinu upotrebe pojedinih izraza i drugih saznanja koja su korisna za kodiranje automatizovanog sistema za prepoznavanje sentimenta u sadržajima pisanim na srpskom jeziku.

#### 4.1.3.3. Statistička eksploracija i prikaz anotiranog korpusa

Osnovni izvor podataka za istraživanje koje je predmet doktorske disertacije predstavlja sajt za recenziranje nastavnog osoblja sa visokoškolskih institucija. Recenzije u svojoj osnovi predstavljaju kritičku ocenu pojedinca i, kao takve, subjektivne su prirode. U situacijama kada se u jednoj rečenici iskazuje više različitih osećanja o različitim aspektima, jednoj rečenici se može pridružiti više anotacija, otuda je broj načinjenih anotacija veći od ukupnog broja rečenica u korpusu – 8101. U 7992 slučaja (98.7%) identifikovan je sentiment, što potvrđuje pretpostavku da korisnici sajta *Oceni profesora* iskazuju osećanja u svojim komentarima. Kako mali procenat korpusa sadrži činjenice, analiza subjektivnosti se nije radila. Tabela 4.12. daje prikaz anotacija prema polaritetu sentimenta i njihovu distribuciju u skupu za obučavanje i testiranje.

Tabela 4.12. Pregled anotacija prema polaritetu sentimenta

	Skup za obučavanje	Skup za testiranje	Ukupno
<b>Pozitivan</b>	3490	1451	4941
<b>Negativan</b>	2117	934	3051
<b>Neutralan</b>	87	22	109
<b>Ukupno</b>	5694	2407	8101

Podaci ilustruju prisustvo asimetrije u distribuciji sentimenta u anotacijama u korist pozitivnog. U odnosu na distribuciju polarizovanog sadržaja na nivou dokumenta, gde je identifikovano 2.5 puta više pozitivnih sadržaja nego negativnih, ovakav disbalans je smanjen na odnos 1.5:1, što će značajno smanjiti probleme obučavanja algoritma klasifikacije nad nebalansiranim podacima. Kako je u skupu zanemarljivo prisustvo neutralnih sadržaja, može se zaključiti da je obuhvatanje ovog polariteta klasifikatorom sentimenta manje relevantno jer je velika većina polarizovanog sadržaja pozitivna ili negativna. Ovako visok procenat sadržaja obojenog sentimentom ukazuje na to da bi klasifikacija aspekata bez identifikovanja sentimenta bila takođe vredna analiza studentskog mišljenja. Naime, kako je u skupu u 98.7% slučajeva prisutan pozitivan ili negativan sentiment, sistem koji bi kategorizovao rečenice prema aspektu bi na koristan način struktuirao recenzije studenata obuhvatajući ujedno i sentiment.

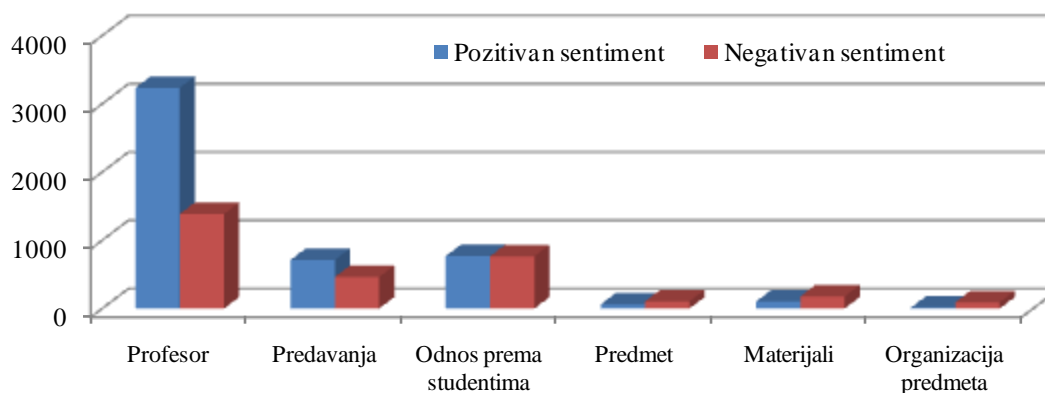
Kako tabela 4.13. pokazuje, u svega 144 slučaja sadržaju je pridružen *Drugi aspekt*, što ukazuje na 98.22% sadržaja iz domena visokog obrazovanja. Ovako visok procenat sadržaja iz domena koji je predmet analize govori da su studenti izuzetno fokusirani u pisanju recenzija i da je većina dostupnih informacija na sajtu *Oceni profesora* relevantna po pitanju profesora koji je predmet pisanja ili jednog od aspekata predefinisanih anotacionom shemom. Na osnovu podataka iz tabele 4.13. može se zaključiti da je dominantni predmet pisanja aspekt *Profesor*. S obzirom na prirodu sajta *Oceni profesora* ovakva distribucija je bila očekivana.

Tabela 4.13. Pregled anotacija prema vrednostima atributa Aspekt

	Skup za obučavanje	Skup za testiranje	Ukupno
<b>Profesor</b>	3348	1321	4669
<b>Predavanja</b>	821	365	1186
<b>Odnos prema studentima</b>	1005	533	1538
<b>Predmet</b>	133	43	176
<b>Materijali</b>	202	78	280
<b>Organizacija predmeta</b>	74	34	108
<b>Drugi aspekt</b>	111	33	144
<b>Ukupno</b>	5694	2407	8101

Slika 4.12. daje prikaz distribucije pozitivnih i negativnih komentara u celom skupu, prema aspektima. Kako je u skupu zanemarljivo prisustvo neutralnih komentara, oni su izuzeti sa

ilustracije, kao i anotacije *Drugog aspekta*, jer sadržaji kojima je ova vrednost dodeljena ne pripadaju domenu visokog obrazovanja. Na osnovu slike može se zaključiti da studenti u većini komentara pozitivno pišu o profesorima i predavanjima, dok se predmeti, nastavni materijali i organizacija predmeta više kritikuju nego hvale.



Slika 4.12. Distribucija pozitivnih i negativnih komentara prema aspektima

Iz tabele 4.14. može se zaključiti da korpusom dominira snažan sentiment, koji čini 65% identifikovanog sentimenta, što može ukazati na čvrst stav koji studenti zauzimaju u svojim recenzijama.

Tabela 4.14. Distribucija anotacija prema intenzitetu iskazanog sentimenta

	Skup za obučavanje	Skup za testiranje	Ukupno
<b>3 - jak</b>	3499	1698	5197
<b>2 - srednji</b>	1596	573	2169
<b>1 - slab</b>	512	114	626
<b>Ukupno</b>	5607	2385	7992

Na nivou izraza anotacija je rađena za sentiment reči i negaciju. Tabela 4.15. sumira statistike identifikovanih sentiment izraza. U korpusu je ukupno identifikovano 8176 pozitivnih i 4555 negativnih sentiment reči i fraza. U skupu za obučavanje je označeno 1258 jedinstvenih pozitivnih reči i fraza i 1521 negativnih, dok je u skupu za testiranje 816 pozitivnih i 923 negativnih reči i fraza. U [Prilogu F](#) prikazani su jedinstveni oblici pozitivnih i negativnih sentiment reči koje su anotatori tokom anotacije označili. Iz ovoga se može zaključiti da studenti koriste bogatiji rečnik u iskazivanju svog nezadovoljstva te da ga detaljnije opisuju. U prilog ovome ide i činjenica da studenti koji uz komentare ostavljaju ocenu 1 za predefinisane aspekte u proseku ostavljaju duže komentare od onih koji ocenjuju predefinisane aspekte ocenom 5. U proseku je po dokumentu (recenziji) izdvojeno 2.12 pozitivnih, odnosno 1.18 negativnih reči, u zavisnosti od sentimenta koji komentar nosi. Ukoliko se izuzmu komentari u kojima nijedna

pozitivna reč nije izdvojena, u 2873 komentara je identifikovano 1 ili više pozitivnih izraza, a u proseku 2.85 pozitivnih izraza po komentaru, dok je u 1658 komentara izdvojeno 1 ili više negativnih izraza, a u proseku 2.75. Ovo ukazuje na činjenicu da se u pojedinim komentarima pojavio mešoviti sentiment. Takva su 703 komentara u korpusu, što čini 18.20% korpusa. Isključivo pozitivni komentari čine 56.17% (2170) korpusa, isključivo negativni komentari čine 24.72% (955), dok ostatak čine komentari u kojima nije izdvojena nijedna sentiment reč.

**Tabela 4.15. Statistička analiza sentiment izraza u korpusu**

Statistike	Pozitivna reč	Negativna reč
Ukupan broj sentiment reči	8176	4555
Broj jedinstvenih sentiment reči	2074	2444
Minimalan broj izraza po dokumentu	0	0
Maksimalan broj izraza po dokumentu	20	30
Prosečan broj izraza po dokumentu	2.12	1.18

Slike 4.13. i 4.14. ilustruju oblake reči polarizovanih sentiment izraza kreiranih u Voyant alatu<sup>40</sup>. Oblak reči prikazuje frekvenciju termina koji se pojavljuju u korpusu. Termini koji se javljaju češće prikazani su u većem formatu. Alat tretira svaku reč odvojenu razmakom kao zaseban termin. U negativnim sentiment izrazima negacija se učestalo koristi. Kako bi se identifikovani negativni sentiment izrazi zadržali u kombinaciji sa negacijom, ključne reči negacije spojene su sa terminom koji negiraju. Kako slike ilustruju, korpusom dominiraju reči dobar, korektan(a), odličan, najbolji, pohvale, dok se korpusom prostiru uglavnom sledeći negativni izrazi: dosadna, loš(e), ne zna, napamet, sramota, šteta, nekorektan.



Slika 4.13. Oblak reči pozitivnih sentiment izraza

<sup>40</sup> Dostupan na stranici <http://voyant-tools.org/>.

Slika 4.14. Oblak reči negativnih sentiment izraza<sup>41</sup>

Tabela 4.16. daje prikaz deset najfrekventnijih pozitivnih i negativnih sentiment reči u korpusu. Kako se može primetiti, pozitivne reči se doslednije upotrebljavaju, te se češće ponavljaju u korpusu, dok su negativni izrazi raznovrsniji i imaju slabiju frekvenciju. Kompletna lista jedinstvenih pozitivnih i negativnih reči može se naći u [Prilogu F](#). Sentiment izrazi su pretežno pridevi. Otuda se može zaključiti da su pridevi važni indikatori osećanja, te bi njihovo pronalaženje u analiziranim sadržajima na osnovu POS tagova predstavljalo bitan korak, što je detaljnije objašnjeno u [Poglavlju 3.2.2](#). Međutim, za srpski jezik ne postoje javno dostupni alati za automatizovano označavanje vrsta reči, odnosno *POS tagging*. Dizajn i implementacija takvog alata zahtevaju interdisciplinarni pristup i prevazilaze okvire doktorske disertacije.

Tabela 4.16. Deset najučestalijih pozitivnih i negativnih reči

Lema	Učestalost	Lema	Učestalost
korektan	400	loš	180
odličan	393	ne zna	74
najbolji	374	dosadan	67
dobar	353	problem	32
pohvala	189	ne može	30
trudi	104	nula	30
pomogne	103	napamet	28
čast	101	šteta	27
super	91	sramota	27
susret	90	katastrofa	26

<sup>41</sup> Negacija je spojena sa izrazom koji negira kako bi Voyant alat tretirao složenicu kao jednu reč.

Iako nose jasan sveukupni sentiment, u pojedinim rečenicama nije bilo moguće identifikovati eksplicitan sentiment izraz. Naredni primeri ilustruju ove slučajeve:

a) *Gde nam njega staviste da predaje.*

U gorenavedenoj rečenici student jasno iskazuje svoje nezadovoljstvo predavačem, međutim, ne koristi eksplicitan izraz.

b) Pokušala sam da nađem vezu kod nje i nisam uspela. *Odbila je bukvalno najrođenijeg.*  
Svaka čast, gospoja!

U gorenavedenom komentaru student opisuje profesora koji je pošten, što student ceni. Međutim, reč odbila sama po sebi nosi negativan sentiment, dok u kontekstu u kojem je upotrebljena u ovom komentaru ukazuje na pozitivan sentiment, te se ona nije izdvojila kao pozitivna reč.

c) *Ja njoj ne bih dao cveće da zaliva, a ona predaje na fakultetu.* Mislim da je i ona svesna svog neznanja.

U gorenavedenom primeru student opisuje nesposobnost profesora, međutim, ne koristi nijednu eksplicitnu sentiment reč kojom iskazuje to svoje nezadovoljstvo. Nezadovoljstvo je skriveno u kontekstu i konstrukciji rečenice.

d) Dobro sprema vežbe i *nema milosti prema zabušantima.*

Izolovano posmatrano reči *nema milosti* nose negativan sentiment, međutim, u navedenoj rečenici student iskazuje svoje zadovoljstvo ovakvim stavom profesora prema studentima.

e) Stručna i veoma posvećena svom poslu. *Ne postoji student koji je ne voli.*

Student piše o profesoru kojeg svi studenti vole, međutim, reči *ne postoji* i *ne voli* same po sebi nose negativan sentiment. Stavljene u kontekst, one govore o obožavanom profesoru.

Rečenice u kojima se na implicitan način iskazuje sentiment ukazuju na potrebu za primenom sentiment analize upotrebom metoda i tehnika mašinskog učenja kako se sentiment skriven u komentarima studenata ne može identifikovati samo na osnovu rečnika pozitivnih i negativnih pojmova. Postoje određene konstrukcije rečenica, kao i fraze, koje se koriste, a koje algoritmi moraju naučiti kako bi efikasno prepoznavali sentiment u sadržajima.



Korpus je sačinjen pretežno iz kratkih komentara. Emotikoni se često koriste, samostalno ili kao dopuna šireg komentara. Posmatrani samostalno, oni nose jasnu poruku: pozitivan emotikon nosi pozitivan sentiment, negativan emotikon nosi negativan sentiment. Međutim, pozitivni emotikoni se ne upotrebljavaju isključivo u pozitivnim komentarima. Vrlo često se koriste kako bi ublažili negativan komentar ili pojačali ironiju iskazanu u komentaru. Nekonzistentnost u njihovoj upotrebi govori da se emotikoni ne mogu upotrebiti kao atributi za klasifikaciju sentimenta. Intenzitetu sentimenta doprinosi znak uzvika, koji se javlja kako na kraju pozitivno obojenih komentara, tako i na kraju negativnih.

Tokom anotacije, anotatori su vodili beleške<sup>42</sup>, te je na osnovu njihovih zapažanja izdvojena lista intenzifikatora. Intenzifikatori doprinose jačanju intenziteta sentimenta, ali dvosmerno. Ukoliko je sentiment komentara ocenjen kao pozitivan, intenzifikator će ga pojačati u pozitivnom ili negativnom smeru, kao i u slučaju negativnih komentara: intenzifikator ga može učiniti izrazitije negativnim, ali mu može i pojačati pozitivan aspekt, što se uočava smanjenjem intenziteta. Nad celim skupom je identifikovano 95 različitih intenzifikatora koji povećavaju ili umanjuju iskazani sentiment, [Prilog G](#). U funkciji intenzifikatora sentimenta javljaju se prilozima, pridevi, rečice i veznici. Prilozima se javljaju uz prideve i druge priloge i određuju stepen osobine koja je iskazana. Najčešće se koriste *prilozima za količinu* – malo, pomalo, mnogo, prilično, suviše, dosta, puno, sasvim, totalno, maksimalno, enormno itd.; *prilozima za način* – jako, ubedljivo, tako, pažljivo, lako, brzo, sporo itd. i *pridevi* – visoki, pravi, potpuni itd. Oni se kombinuju sa pozitivnim ili negativnim rečima i jačaju ili slabe intenzitet iskazanog sentimenta. Pored priloga za količinu u funkciji iskazivanja količine ili intenziteta javlja se konstrukcija sa komparativom priloga malo/mного i veznikom *nego* (na primer: Predavanja su više zanimljiva nego korisna).

Pojedini intenzifikatori ne samo da pojačavaju sentiment nego mogu i promeniti kontekst, u zavisnosti od upotrebe (primer: blago). Intenzifikatori poput *sve u svemu* ukazuju na negativan sentiment ili umanjuju iskazani sentiment, nikada ga ne jačaju. Rečice takođe doprinose intenzitetu sentimenta. Osim odrične rečice *ne*, najčešće se javljaju *modalne rečice* (one koje ističu lični stav): *zaista*, *verovatno*, *valjda*, i sl. Tabela 4.17. ilustruje primere pojedinih intenzifikatora i njihove upotrebe.

Najfrekventniji veznici su suprotni i sastavni. Suprotni veznici, u najvećem broju slučajeva *a*, *ali*, *no*, *nego* i *već*, dovode do promene sentimenta u komentaru. Sa stanovišta lingvistike ovo ne iznenađuje, s obzirom na to da navedeni veznici iskazuju suprotni odnos – odnos kojim se povezuju rečenice među čijim sadržajima postoji nepodudarnost ili suprotnost (Stanojčić, et al., 2002). Veznik *a* pojačava intenzitet sentimenta u predikatskoj rečenici koja sledi, odnosno u delu komentara u kome se menja aspekt ili intenzitet. Upotrebljen samostalno veznik *a* posebno ističe deo komentara koji će uslediti, dok u kombinaciji sa veznikom *i* ističe ono što je ranije rečeno. Veznik *ali* nije pojačivač te snage. On se često javlja između udvojenih priloga, poput *nikada*, *ali*

<sup>42</sup> Sa aspekta lingvistike i gramatike izdvojena zapažanja anotatora interpretirala je Gordana Svilengaćin, master profesor jezika i književnosti (srbista).

*nikada; jako, ali jako; veoma, ali veoma; izuzetno, ali izuzetno.* U ovom slučaju, veznik *ali* ne funkcioniše kao suprotni veznik, već kao rečca za posebno isticanje. Među nezavisnim rečenicama veznik *ali* označava i nekoliko pododnosa. Jedan od njih je koncesivni. Koncesivnim odnosom se pokazuje da se na osnovu sadržaja prve rečenice ne očekuje sadržaj druge rečenice (Stanojčić, et al., 2002). Ovaj odnos postoji i među zavisnim rečenicama, ali tada se iskazuje veznicima *mada, premda, iako*. Navedenim veznicima se sentiment u komentarima menja iz pozitivnog u negativan i obrnuto. Od sastavnih veznika najčešće se koristi veznik *i*. Njime se sentiment ne menja, nego se stav studenta dodatno argumentuje. Za pohvalu ili pokudu koriste se i konstrukcije koje se zbog česte upotrebe u nepromenjenom obliku mogu svrstati u fraze: *bila mi je čast, jedan od najboljih, komentar je suvišan* itd. U [Prilogu H](#) je data lista identifikovanih domensko specifičnih fraza.

**Tabela 4.17. Intenzifikatori sentimenta i njihova upotreba**

Intenzifikator	Primer upotrebe	Anotacija
veoma	Veoma ljubazna.	Pozitivan sentiment, intenzitet 3.
	Ljubazna.	Pozitivan sentiment, intenzitet 2.
	Veoma loš profesor.	Negativan sentiment, intenzitet 3.
	Loš profesor.	Negativan sentiment, intenzitet 2.
krajnje	Krajnje profesionalno pristupa predavanjima.	Pozitivan sentiment, intenzitet 3.
	Krajnja arogancija.	Negativan sentiment, intenzitet 3.
blago	Blago FONa.	Pozitivan sentiment, intenzitet 3.
	Blago rečeno neprijatan.	Negativan sentiment, intenzitet 3.
	Blago fakultetu koji je ima i studentima koji je slušaju.	Pozitivan sentiment, intenzitet 3.
malo	Malo dosadna predavanja.	Negativan sentiment, intenzitet 1.
	Dosadna predavanja.	Negativan sentiment, intenzitet 2.
retko	Retko dobar predavač.	Pozitivan sentiment, intenzitet 3.
	Retko loš čovek.	Negativan sentiment, intenzitet 3.

U korpusu je identifikovano 1720 negacija (1176 u skupu za obučavanje i 544 u skupu za testiranje) i 31 jedinstveni oblik negacije prikazan u [Prilogu I](#). Rečca *ne* se najčešće pojavljuje u korpusu, tabela 4.18. Ona se koristi uz glagole. Ukoliko se javi pre glagola pozitivne semantike, ova rečca doprinosi negativnom sentimentu rečenice, i obratno – *ne* uz glagol negativne semantike daje pozitivan komentar.

Tabela 4.18. Najčešći oblici negacije

Oblik negacije	Broj pojavljivanja
ne	793
nije	290
nema	124
nisu	85
ni	80

Prema uputstvu za anotaciju, anotatori su označavali negaciju glagola. Međutim, da bi se na adekvatan način modelovala upotreba negacije, bilo je potrebno označavati i odrične zamenice i priloge (niko, ništa, nikad, nigde, ničiji i sl.). Tako se i veznici *ni* i *niti* koriste u funkciji negiranja. Kada se u rečenici upotrebe odrične zamenice i priloge, tada se koristi i takozvana dupla negacija, odnosno dvostruko negiranje. Ukoliko se želi pokazati da se sadržaj komentara ne odnosi ni na jedan slučaj, nije dovoljno samo da se glagolu doda odrična zamenica ili prilog (ništa, nijedno, nikada i dr.), već se i sam glagol mora staviti u odričan oblik (Stanojčić, et al., 2002). U komentaru sa dvostrukom negacijom najmanje jedan glagol mora biti u odričnom obliku, a kako bi forma dvostruke negacije bila zadovoljena, tom glagolu biće dodat drugi glagol, takođe odrični, ili odrična zamenica ili prilog. U takvim situacijama negacije se ne poništavaju međusobno. Primer: *On ništa ne zna; Ne postoji [nijedan] student koji je ne voli.*

Dupla negacija se učestalo pojavljuje u korpusu i nosi sa sobom određene probleme. Naime, ukoliko se u rečenici dva puta negira reč pozitivne semantike, rečenica može nositi i negativni sentiment (primer *Ne mogu da dam nijednu pozitivnu ocenu*), ali i pozitivan (primer *Ne postoji student koji je ne voli*). To navodi na zaključak da se ne može izvesti jedinstveno pravilo po pitanju načina upotrebe duple negacije. Naredni primeri ilustruju dodatnu problematiku:

- a) *Nemam nijednu zamerku.*
- b) *Nemam nijednu pozitivnu reč.*

U navedenim rečenicama koristi se odrična zamenica *nijedan* i negacija *nemam*, sentiment rečenice se menja u zavisnosti od toga da li se one kombinuju sa rečima pozitivne ili negativne semantike. Ovakva upotreba negacije i odrične zamenice dosledna je u korpusu. Međutim, drugačije konstrukcije rečenica (naredni primer) predstavljaju izuzetak od ovog pravila.

- c) *Nijedno drugo predavanje nije toliko raznoliko i zanimljivo.*

U ovim rečenicama, iako se negiraju reči pozitivne semantike, konačni sentiment rečenice je pozitivan jer student u konkretnom primeru na ovaj način ističe da su predavanja profesora kojeg ocenjuje raznolika i zanimljiva, za razliku od drugih predavanja.

- d) *Ma žena nema šta ne zna, kapa dole.*  
e) *A vala ni ona teorija, nikad lakše nešto na FONu nisam naučila i sa većim elanom.*  
f) *Da ne kasni i da pokazuje više energije na časovima ne bi bilo zamerki.*

U prvom primeru glagol *nema* negira neznanje profesora, odnosno poništava narednu negaciju, dok se u drugom primeru negacije ne poništavaju. U poslednjem primeru se dva puta negiraju reči negativne semantike i celokupna rečenica nosi negativan sentiment.

Gorenavedeni primeri potvrđuju činjenicu da se jedinstvenim pravilima ne može modelovati višestruka negacija, koja je učestala u srpskom jeziku. Generalno posmatrano, možemo zaključiti da na sentiment komentara samo negiranje ne utiče toliko koliko ona činjenica koja se negira, a koja je iskazana imenicom, pridevom ili drugim glagolom i sama po sebi nosi pozitivnu ili negativnu konotaciju.

Prestanak važenja negacije se javlja kod interpunkcijskih znakova, zagrada ili pojave druge negacije. Međutim, i određeni veznici dovode do prestanka važenja negacije (poput već, jer, ako, samo, ili, iako i dr.). Pojedini od tih veznika menjaju sentiment, pojedini pojačavaju iskazani sentiment u drugom delu rečenice, ili pak potvrđuju i objašnjavaju prethodno iskazani sentiment.

Tokom anotacije identifikovane su i sledeće specifičnosti u konstrukciji rečenica. Kako bi se iskazao pozitivan komentar izrazitog intenziteta, koriste se zavisne kondicionalne rečenice. Njima se označava uslov za realizovanje situacije označene višom rečenicom, (Stanojčić, et al., 2002). Primeri njihove upotrebe navedeni su u nastavku:

- a) *Da je više mladih profesora na katedri za računarstvo, koji su u koraku sa vremenom, a ne oni koje je vreme davno davno pregazilo....*  
b) *Da nije bilo nje, bila bi totalna katastrofa.*

U komentarima su uočeni različiti stepeni poređenja prideva i priloga (onih koji su po porecku pridevi): pozitiv, komparativ, superlativ. Kada je upotrebljen pozitiv (primer: dobar, loš) ili superlativ (primer: najbolji, najgori), tada je sentiment nedvosmislen:

pozitiv prideva pozitivne semantike = pozitivan sentiment

superlativ prideva negativne semantike = negativan sentiment

Kombinacija modalnih glagola *moći*, *trebati* i komparativa prideva i priloga daje negativan sentiment:

a) Sve **može** mnogo **bolje**.

b) Knjige su dobro napisane, opet **može** to mnogo **bolje** tj. na mnogo manje strana da se odradi, a ne da se mlati prazna slama..

c) Literatura bi **trebalo** da bude malo **bolje** odrađena.

Izuzetak od ovog pravila je kada se reč *bolji* (komparativ prideva dobar) kombinuje sa rečima pozitivnog sentimenta. Tada doprinosi komentaru izrazito pozitivnog sentimenta.

*Dobar je profesor, a još **bolji** čovek...*

Obratno, reč negativnog sentimenta i reč *gori* (komparativ prideva loš) doprinose komentaru izrazito negativnog sentimenta.

*Najgori dekan u istoriji fakulteta, loš profesor i još **gori** čovek...*

Pozitiv i superlativ prideva iskazuju pozitivan sentiment, a komparativ negativan.

a) pozitiv prideva = pozitivan sentiment

*Korektan, **dostupan**, savestan i profesionalan!*

b) komparativ prideva = negativan sentiment

*Sve je ok na predavanjima i na vežbama, samo bi profesor mogao malo da bude **dostupniji** studentima i da nam on upisuje ocene sa predispitnih, otkud neko zna koliko sam ja aktivna kod njega na časovima ili koliko puta sam došla na čas..*

Modalni glagoli *moći*, *trebati* se upotrebljavaju i za iskazivanje preporuka u komentarima kada se ovi glagoli koriste u potencijalu (trebalo bi, moglo bi). Potencijal je glagolski oblik (način) sa osnovnim modalnim značenjem i njime se označava stav govornika prema nerealizovanoj radnji (Stanojčić, et al., 2002). Upotreba potencijala uglavnom nosi negativan sentiment jer se komentarima tog tipa ukazuje na propuste profesora.

a) **Trebalo bi** da bude više angažovan jer studenti imaju šta da nauče od njega!

b) To **bi trebalo** malo da revidira i da rastereti slajdove, jer je jako teško pratiti i održati koncentraciju na časovima kada iz slajda u slajd vidite sve više sitno zbijenog teksta i naprežete oči da sve ispratite.

- c) *Ukoliko niste aktivni i ne dolazite redovno na nastavu, po mom mišljenju, to **ne bi trebalo** da utiče drastično na konačnu ocenu.*
- d) *Predavanja su korisna, jasna, i profesorka je uvek voljna da nešto objasni ili pojasni, ali sve **bi moglo** biti mnogo interesantnije, sa drukčijim vidom prezentacije.*
- e) ***Mogao bi** malo više fer da ocenjuje, ne razumem koji mu je kriterijum.*
- f) ***Mogla bi** malo više da se angažuje oko objašnjavanja detalja na časovima vežbi.*

Upotreba zamenica takođe ukazuje na sentiment. Pokazne zamenice *ovaj* i *ova* nose negativan sentiment. Međutim, ne može se definisati apsolutno pravilo kako za oba oblika postoje brojna odstupanja (Neko **ovaj** ispit položi da ne pogleda knjigu. **Ova** profesorka zaslužuje višu ocenu.).

- a) *I **ova** je preko noći došla na fakultet.*
- b) ***Ova** je najluđa od svih.*
- c) ***Ova** je sve završila u Mitrovici.*
- d) ***Ovaj** čovek je jedan od onih veličina...*

Nasuprot ovome, prisvojne znamenice *moj*, *naš* nose pozitivan sentiment.

- a) ***Naš** cenjeni i voljeni dekan.*
- b) *Mogao bi on to puno bolje, ali **naš** je pa ga volimo :).*
- c) *Profa **moj** najbolji na svetu!*

Zapažanja izvedena tokom anotacije upotrebiće se u definisanju pravila automatizovanog sistema za prepoznavanje sentimenta. Očekuje se da će doprineti efikasnosti njegovog rada. Anotacija sentiment reči, negacije i njenog opsega omogućila je kreiranje neophodnih rečnika i njihovu upotrebu u sentiment analizi. Dizajn sistema za identifikaciju negacije, kao ni detaljnije modelovanje negacije u sadržajima na srpskom jeziku nisu predmet doktorske disertacije. Sarkazam i ironija su retko prisutni u korpusu, te se ovim pitanjima doktorska disertacija ne bavi. U narednom poglavlju opisano je formiranje rečnika pojmova, koji ujedno predstavlja poslednji pripremni korak za realizaciju sentiment analize nad kreiranim korpusom.

#### 4.1.4. Formiranje rečnika sentiment pojmova

Reči i fraze koje iskazuju i prenose pozitivan ili negativan sentiment predstavljaju osnovni instrument sentiment analize. Pozitivne sentiment reči se koriste kako bi se iskazalo željno stanje ili kvalitete, dok se negativne koriste za iskazivanje neželjenih stanja ili nepoželjnih karakteristika. Primeri pozitivnih sentiment reči su *korektan*, *dobar*, *odličan*. Primeri negativnih sentiment reči su *loš*, *problem*, *dosadan*. Pored pojedinačnih reči, sentiment može biti iskazan i frazama ili kombinacijom više reči, npr. *brzinom munje*, *ne zna*. Sentiment leksikon objedinjuje pojedinačne sentiment reči, fraze i idiome<sup>43</sup>. Sentiment reči se mogu podeliti na *osnovne* i *komparativne* (Liu, 2012). Prethodni primeri spadaju u grupu osnovnih sentiment reči, dok bi komparativne reči iskazivale određeno poređenje ili mišljenje u komparativu ili superlativu poput: gori, bolji, najbolji.

Pristupi formiranju sentiment rečnika detaljnije su opisani u [Poglavlju 3.3.3](#). U istraživanju koje je predmet doktorske disertacije upotreba uopštenih rečnika sentiment pojmova nije dovoljna usled specifičnosti domena i konteksta. Za potrebe analize sentimenta u sadržajima o visokom obrazovanju kreirano je 5 rečnika na osnovu korpusa i realizovanih anotacija: rečnik pozitivnih sentiment reči, rečnik negativnih sentiment reči, rečnik intenzifikatora, rečnik neutralizatora i rečnik negacije i drugih reči koje menjaju sentiment polaritet naredne reči. Jedinstveni oblici anotiranih pozitivnih i negativnih reči u skupu za obučavanje, koji su prikazani u [Prilogu F.1.](#) i [Prilogu F.2](#), upotrebljeni su za kreiranje domen-specifičnih rečnika sentiment izraza. Uz svaki anotirani izraz pridružena je oznaka sentiment polariteta. Rečnici su prošireni identifikovanim domen-specifičnim frazama prikazanim u [Prilogu H](#). Fraze kojima se iskazuje pozitivan stav pridružene su rečniku pozitivnih sentiment reči, dok su fraze kojima se iskazuje negativan stav pridružene rečniku negativnih sentiment reči. Tabela 4.19. daje prikaz forme ovih rečnika.

**Tabela 4.19. Forma domen-specifičnih sentiment rečnika**

Rečnik pozitivnih sentiment reči	Rečnik negativnih sentiment reči
najbolji: [pozitivno]	bezobrazna: [negativno]
najduhovitiji: [pozitivno]	bije: [negativno]
odličan: [pozitivno]	blati: [negativno]
odlično: [pozitivno]	bojažljiv: [negativno]
nije škrt: [pozitivno]	bruka: [negativno]

Oba rečnika su manuelno ispitana kako bi se uklonile reči koje su anotatori označili kao sentiment reči, a koje ne iskazuju sentiment, zatim reči koje su greškom anotirane suprotnim polaritetom, kao i negacije, koje su uvrštene u zaseban rečnik. Tabele 20. i 21. daju prikaz reči uklonjenih iz rečnika i razloge za njihovo uklanjanje.

<sup>43</sup> U nastavku teksta *Sentiment reč* će se odnositi kako na pojedinačne reči tako i na fraze i idiome.

**Tabela 4.20. Reči uklonjene iz rečnika pozitivnih sentiment pojmova**

Pozitivne sentiment reči	Razlog korekcije
dodir	neutralan sentiment
dosadan	negativan sentiment
duge	neutralan sentiment
lišti	neutralan sentiment
komunikaciju	neutralan sentiment
kolega	neutralan sentiment
konsultacija	neutralan sentiment
maksimalna	modifikator
maksimalni	modifikator
merilima	neutralan sentiment
ne voli	negativan sentiment
oličenje	neutralan sentiment
opseg	neutralan sentiment
predavao	neutralan sentiment
sujetan	negativan sentiment
tu	neutralan sentiment
užas	negativan sentiment
veći	modifikator
većim	modifikator
viši	modifikator
visoke	modifikator
višu	modifikator
nula	negativan sentiment
više	modifikator
pedagog	neutralan sentiment
ima	neutralan sentiment

**Tabela 4.21. Reči uklonjene iz rečnika negativnih sentiment pojmova**

Negativne sentiment reči	Razlog korekcije
bez	modifikator
kurs	neutralan sentiment
mного	modifikator
može biti	modifikator
napiše	neutralan sentiment
nije loš	pozitivan sentiment
pošalje	neutralan sentiment
posle	neutralan sentiment
pravi	neutralan sentiment
puno	modifikator
puno bolje	modifikator
ranije	neutralan sentiment
široko	neutralan sentiment
viši	neutralan sentiment

Na osnovu identifikovanih oblika negacije prikazanih u [Prilogu I](#) kreiran je rečnik negacije, čija je funkcija promena orijentacije sentimenta naredne reči u tekstu. Na osnovu identifikovanih specifičnosti u stilu pisanja i načinu na koji se rečenice konstruišu lista identifikovanih negacija proširena je rečima koje obično menjaju iskazani sentiment, poput *ali*, *mada*. Uz svaku reč je pridružena oznaka *inv*, koja ukazuje na to da reč pripada rečniku negacije i da je njena funkcija promena sentiment polariteta susedne reči. Tabela 4.22. prikazuje formu rečnika negacije.

**Tabela 4.22. Struktura rečnika negacije**

Rečnik negacije
ali: [inv]
mada: [inv]
ne: [inv]
nije: [inv]
nema: [inv]



Kako bi se definisala jačina iskazanog sentimenta, kreirana su dva dodatna rečnika koja omogućavaju identifikovanje polariteta zavisno od specifičnog konteksta u kom su određeni izrazi upotrebljeni, odnosno intenzitet pozitivnosti ili negativnosti. Tokom anotacije anotatori su identifikovali intenzifikatore sentimenta koji, zavisno od tipa, pojačavaju ili umanjuju iskazani sentiment. Intenzifikatori sentimenta, prikazani u [Prilogu G](#), manuelno su razdvojeni u dva rečnika na osnovu njihove funkcije. U prvom rečniku se nalaze reči koje pojačavaju iskazani sentiment i njima je pridružena oznaka *inc*, koja ukazuje na to da pojava date reči u analiziranoj rečenici uvećava intenzitet iskazanog sentimenta. U drugom rečniku se nalaze reči koje umanjuju iskazani sentiment i njima je pridružena oznaka *dec*, koja govori da pojava date reči u analiziranoj rečenici smanjuje intenzitet iskazanog sentimenta. Tabela 4.23. daje prikaz forme ovih rečnika.

**Tabela 4.23. Struktura rečnika intenzifikatora i neutralizatora**

Rečnik intenzifikatora	Rečnik neutralizatora
baš: [inc]	malo: [dec]
krajnje: [inc]	pomalo: [dec]
još kako: [inc]	u suštini: [dec]
veoma: [inc]	sve u svemu: [dec]
previše: [inc]	donekle: [dec]

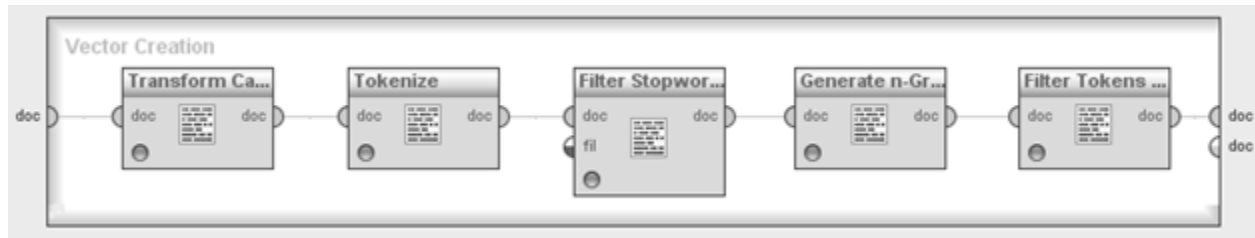
#### 4.1.5. Otkrivanje preovlađujućeg sentimenta primenom inteligentnih metoda i tehnika

Otkrivanje sentimenta skrivenog u studentskim komentarima o visokoškolskim institucijama rađeno je primenom metoda i tehnika nadgledanog učenja i primenom rečnika sentiment pojmova generisanih tokom anotacije korpusa. Naredna poglavlja detaljnije opisuju razvijene modele.

##### 4.1.5.1. Algoritmi mašinskog učenja korišćeni za otkrivanje sentimenta

Primena metoda i tehnika nadgledanog učenja zahteva pripremljen veliki skup poznatih primera za obučavanje, kojima se klasifikator sentimenta obučava kako bi napravio distinkciju između pozitivnih i negativnih komentara, što je omogućeno manuelnom anotacijom prikupljenih podataka o visokoškolskim institucijama i profesorima. U analizi kreiranog korpusa primenjena su tri algoritma nadglednog učenja koja dominiraju u praksi: *Naive Bayes*, *Support vector machines* i *k-Nearest Neighbor*, koji su detaljnije opisani u [Poglavlju 3.3.1](#). Klasifikacija je izvršena na nivou celog komentara i na nivou pojedinačnih rečenica. Tokom anotacije identifikovano je da mali procenat korpusa čine neutralni komentari (1.5% u skupu za obučavanje i 0.9% u skupu za testiranje), te su izgrađeni binarni modeli klasifikacije koji

razlikuju pozitivnu i negativnu klasu sentimenta. Proces preprocesiranja podataka iz fajlova prikazan je na slici 4.15.



Slika 4.15. Proces preprocesiranja podataka<sup>44</sup>

Preprocesiranje podataka u razvijenim modelima podrazumeva:

- Konverziju velikih slova u mala (proces *Transforme Cases*), kojom je omogućeno da se npr. reč „Profesor” tretira isto kao i reč „profesor” i da se pojava navedenih termina broji kao jedinstvena.
- Tokenizaciju sadržaja (proces *Tokenize*), kojom se vrši prelamanje sadržaja dokumenta po rečima kako bi se dobila lista reči. U razvijenim modelima prelamanje se vrši pri pojavi karaktera koji ne predstavljaju slovo, poput crte, praznog mesta i sl.
- Filtriranje stop reči (proces *Filter Stopwords Dictionary*). Stop reči predstavljaju reči koje su učestale u govornom i pisanom jeziku, ali ne donose novo znanje, poput veznika. Razvijeni modeli koriste listu stop reči prikazanu u tabeli 4.24.
- Izdvajanje fraza odnosno n-gramova (proces *Generate n-grams – terms*)<sup>45</sup>. Za svaki razvijeni model eksperimentisano je sa bigramima, trigramima i četirigramima. Svi modeli iskazuju najbolje performanse pri generisanju bigrama, koji povezuju dva tokena, odnosno dve reči u nizu, izuzev SVM klasifikatora celih recenzija, koji ima najbolje performanse kada se koriste samo unigrami, odnosno lista pojedinačnih reči koje se pojavljuju u skupu podataka.
- Filtriranje reči po dužini (proces *Filter tokens by lenght*) kako bi se eliminisale reči koje su sačinjene iz jednog ili dva karaktera, odnosno koje imaju više od 25 karaktera.

Modeli u narednom koraku vrše kreiranje vektora termina primenom *TF-IDF* mere, koja je detaljnije opisana u [Poglavlju 3.2.2](#). Ona pokazuje relativnu značajnost termina za određeni dokument u odnosu na značajnost koju termin ima za sve dokumente u korpusu. Ukoliko se posmatrani termin pojavljuje relativno često u jednom dokumentu, *TF-IDF* mera će biti viša i pokazivaće da data reč karakteriše konkretni dokument. Ukoliko veliki broj dokumenata u korpusu sadrži posmatrani termin, *TF-IDF* mera će biti niža i ukazivaće na neznačajnost date

<sup>44</sup> Proces je kreiran u RapidMiner alatu.

<sup>45</sup> Pod frazama se na ovom mestu podrazumevaju reči koje se frekventno pojavljuju u korpusu jedna za drugom. One su izdvojene na osnovu njihovog učestalog zajedničkog pojavljivanja, te ih je potrebno razlikovati od lingvističkih fraza, koje studenti učestalo koriste u iskazivanju sentimenta (poput: *kapa dole*) i koje su ranije spominjane.

reči. Rezultat preprocesiranja i primene *TF-IDF* mere jeste skup podataka u kojem jedan red predstavlja dokument (jedan komentar ili jedna rečenica), dok kolone predstavljaju reči koje se pojavljuju u celom korpusu. Dodeljene vrednosti ukazuju na značaj koju reč ima za konkretan dokument. Vrednosti bliže jedinici ukazuju na veću značajnost, dok nula pokazuje da se reč ne pojavljuje u analiziranom dokumentu.

Tabela 4.24. Stop reči

a	hoćete	jeste	meni	nego	nisam	njima	s	ta	vama
ako	hoćeš	jesu	mi	neka	nisi	njoj	sa	tada	vas
ali	hoću	joj	mimo	neki	nismo	nju	sam	taj	vaš
bi	i	još	moj	nekog	niste	no	samo	tako	vaša
bih	iako	ju	moja	neku	nisu	o	se	te	vaše
bila	ih	kada	moje	nema	njega	od	sebe	tebe	već
bili	ili	kako	mu	netko	njegov	odmah	sebi	tebi	vi
bilo	iz	kao	na	neće	njegova	on	si	ti	vrlo
bio	ja	koja	nad	nećemo	njegovo	ona	smo	to	za
bismo	je	koje	nakon	nećete	njemu	oni	ste	toj	zar
biste	jedna	koji	nam	nećeš	njezin	ono	su	tome	će
biti	jedne	kojih	nama	neću	njezina	ova	sve	tu	ćemo
da	jedno	kojima	nas	nešto	njezino	pa	svi	tvoj	ćete
do	jer	koju	naš	ni	njih	pak	svog	tvoja	ćeš
duž	jesam	kroz	naša	nije	njihov	po	svoj	tvoje	ću
ga	jesi	li	naše	nikoga	njihova	pod	svoja	u	što
hoće	jesmo	me	našeg	nikoje	njihovo	pored	svoje	uz	šta
hoćemo	jest	mene	ne	nikoju	njim	pre	svom	vam	

Slika 4.16. ilustruje rezultat preprocesiranja tekstova. Komentar prikazan na slici: „*Krajnja arogancija i nipodaštavanje prema studentima*“, najbolje opisuje reč *arogancija*, što se zaključuje na osnovu vrednosti pridružene ovoj reči – 0.744. Na osnovu nultih vrednosti dodeljenih ostalim rečima vidljivim na slici može se zaključiti da one nisu prisutne u konkretnom komentaru. Takođe, slika ilustruje i generisane bigrame. Reči *aktivno* i *učestvuju*, odnosno *bahat* i *stav* se učestalo pojavljuju jedna za drugom u korpusu. Rezultat preprocesiranja je i lista reči sa frekvencijom njihovog pojavljivanja u korpusu, kao i u pozitivnim i negativnim komentarima. Modeli čuvaju reprezentativnu listu reči sa podacima o frekvenciji njihovog pojavljivanja za kasnije upotrebe.

ExampleSet (2575 examples, 5 special attributes, 4024 regular attributes)						View Filter (2575 / 2575): all				
Row No.	text	label	metadata_file	metadata_p...	metadata_d...	aktivno_učestvuju	aktuelan	antipatičan	arogancija	bahat_stav
1878	Krajnja arogancija i nipodaštave	negativno	1603.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0.744	0
2524	Arogancija i visil	Krajnja arogancija i nipodaštavanje prema studentima.					0	0	0.608	0
1985	arogancija mak						0	0	0.460	0
1	Sve super!						0	0	0	0
2	Ana je najjača!	Press "F3" for focus.					0	0	0	0
3	Dobar i korektan profesor fin i sj	pozitivno	1002.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0
4	Dobar i objektivan profesor uvek	pozitivno	1003.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0
5	Profesorka Mina je divna, vrlo je	pozitivno	1004.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0
6	Spreman uvek da pomogne.	pozitivno	1005.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0
7	uvek spremna da pomogne.	pozitivno	1006.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0
8	Veoma dobar profesor spreman	pozitivno	1007.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0
9	veoma posvećen poslu, ume da	pozitivno	1008.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0
10	Blago!	pozitivno	1009.bt	C:\Users\Ko	Sep 2, 2015	0	0	0	0	0

Slika 4.16. Rezultujući skup podataka

Nad pripremljenim skupom podataka primenjuju se algoritmi: *Naïve Bayes*, *Support vector machines* i *k-Nearest Neighbor* i evaluiraju se performanse izgrađenog modela.

### Podешavanje parametara algoritama

U definisanju optimalne kombinacije parametara, za svaki primenjeni algoritam, izvršen je niz eksperimenata u kojima su varirane vrednosti dostupnih parametara. Rezultati rada rezultujućih modela, primenjeni nad podacima iz skupa za obučavanje, su upoređeni, a u nastavku su prikazane vrednosti parametara primenjenih algoritama koji daju najpreciznije rezultate klasifikacije.

#### 1. Naïve Bayes klasifikator

Ukoliko analizirani dokument  $D$  sadrži pojam  $T$  koji se nije pojavio ni u jednom dokumentu klase  $C$  (u skupu za obučavanje), onda će verovatnoća  $P(T/C)$  biti 0 (nula). Problem nulte verovatnoće ogleda se u činjenici da njena pojava poništava sve informacije o ostalim verovatnoćama kada su one pomnožene prema originalnoj Naïve Bayes formuli (Maimon, et al., 2005), koja je prikazana u [Poglavlju 3.3.1.4](#). Da bi se sprečio uticaj nulte verovatnoće koristi se tzv. *Laplace correction*, koji najčešće podrazumeva korigovanje vrednosti verovatnoća dodavanjem svakoj verovatnoći vrednost 1 (Zhang, et al., 2004). Na ovaj način se postiže da svaka reč koja se pojavi u dokumentu ima verovatnoću veću od nule, čime je sprečena pojava nulte verovatnoće. Iz ovog razloga, u podešavanjima *Naïve Bayes* algoritma aktiviran je parametar *Laplace correction*.

#### 2. Support vector machines klasifikator

Centralni parametar *Support vector machines* algoritma je *kernel* funkcija, kojom se računa sličnost između dokumenata. Kako bi se ustanovila optimalna *kernel* funkcija, kreirani su SVM modeli u kojima su upotrebljene dostupne funkcije (*dot*, *radial*, *polynomial*, *neural*, *anova*,

*epachnenikov*, *Gaussian combination* i *multiquadric*). Poređenjem mera koje oslikavaju performantnost rada razvijenih modela ustanovljeno je da SVM model klasifikacije daje najbolje rezultate koristeći *dot kernel* funkciju:

$$K(x, y) = x * y$$

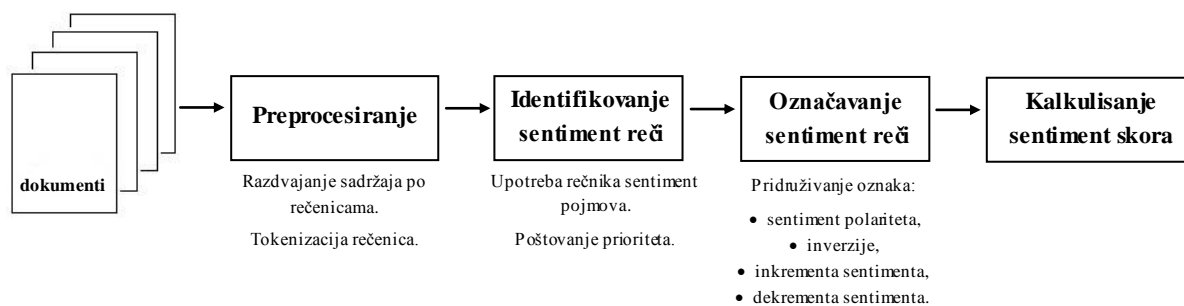
### 3. k-Nearest Neighbor klasifikator

Prvi korak *k-Nearest Neighbor* algoritma jeste pronalaženje k primera za obučavanje koji su najbliži nepoznatim primerima. Optimalna vrednost parametra je  $k=3$ , dok su razvijeni modeli i za  $k=1$  i  $k=5$ . Modeli iskazuju najbolje performanse uz numeričku meru *cosine similarity* za pronalaženje najbližih suseda. Sličnost između dva vektora ili dva dokumenta u vektoru prostora izračunava se na osnovu kosinusa ugla između njih. Može se posmatrati kao poređenje dokumenata u normalizovanom prostoru jer mera ne uzima u obzir samo pojavljivanje reči (*TF-IDF*) u svakom dokumentu, nego ugao između dokumenata na osnovu kojih se procenjuje njihova povezanost i sličnost<sup>46</sup>.

Razvijeni modeli i konfiguracija njihovih procesa prikazana je u *Prilogu J*. Procena preciznosti klasifikacije prikazana je u [Poglavlju 4.3.1](#).

#### 4.1.5.2. Sentiment analiza zasnovana na leksikonima

Otkrivanje polariteta iskazanog mišljenja studenata na bazi leksikona i jednostavnih pravila realizovano je na nivou celog komentara i na nivou pojedinačnih rečenica prema proceduralnim koracima prikazanim na slici 4.17.



Slika 4.17. Proces sentiment analize

Zavisno od nivoa na kojem se vrši sentiment analiza dokumenti predstavljaju celokupne recenzije ili anotirane rečenice prema polaritetu. Da bi se nad dokumentima mogla realizovati analiza, oni moraju zadovoljiti sledeću sturukturu:

<sup>46</sup> Detaljniji opis mere *cosine similarity* dostupan je na sajtu <http://blog.christianperone.com/2013/09/machine-learning-cosine-similarity-for-vector-space-models-part-iii/>.

- Svaki tekst je lista rečenica (kada se radi analiza na nivou cele recenzije).
- Svaka rečenica je lista tokena.
- Svaki token je torka sačinjena iz sledećih elemenata: tačan oblik reči koji se pojavio u tekstu i lista pridruženih obeležja za polaritet, inverziju, inkrement ili dekrement.

Dokumenti predstavljaju ulaz za preprocesiranje, a rezultat je kolekcija rečenica, pri čemu je svaka rečenica ujedno i kolekcija tokena. *Algoritam K.1.* prikazan u [Prilogu K](#) opisuje preduzete korake preprocesiranja dokumenata.

Naredni korak podrazumeva identifikovanje pozitivnih i negativnih sentiment reči u dokumentima na osnovu kreiranih rečnika, opisanih u [Poglavlju 4.1.4](#). Model obezbeđuje da se uz reč ili frazu pridruži isključivo jedna oznaka, a da se pri tom poštuju dva pravila prioriteta:

1. Dužim rečima ili frazama se daje prioritet. Na ovaj način se obezbeđuje da se reč ne označava višestruko, te ukoliko je u jednoj rečenici upotrebljena fraza, da se ona u celini označi umesto označavanja pojedinačnih reči koje je sačinjavaju. Na primer, zbog učestalosti upotrebe, tekst „*može to bolje*“ okarakterisan je kao fraza. Ukoliko se ovo pravilo ne bi poštovalo, reč *bolje* bi bila označena kao reč koja nosi pozitivan sentiment, dok celokupna fraza nosi negativan sentiment.
2. Pretraga se vrši sleva nadesno, čime se postiže redosledna pretraga sadržaja.

Označavanje identifikovanih sentiment reči u dokumentima predstavlja naredni korak čiji rezultat jeste ulazni tekst obogaćen oznakama tipa *pozitivno*, *negativno*, *inv*, *dec*, *inc* (oznake koje su pridružene pojedinim rečima zavisno od rečnika kojem pripadaju i funkcije koju imaju u rečenici – za detaljniji opis pogledati [Poglavlje 4.1.4](#)). Za identifikovanje i označavanje reči iz rečnika upotrebljen je *Algoritam K.2*, prikazan u [Prilogu K](#). Rezultat označavanja prikazan je na slici 4.18. Komentar koji je predmet analize podeljen je na reči, a svakoj reči je pridružena oznaka. Reč „pomalo“ ima funkciju neutralizatora sentimenta i označena je oznakom *dec*, dok reč „dosadna“ predstavlja sentiment reč negativne semantike i pridružena joj je oznaka *negativno*. Nad označenim sadržajem primenjuje se mera sentimenta.

```
Analizirana recenzija studenta:  
Pomalo dosadna.
```

```
Lista svih reci u recenziji  
["['pomalo', 'pomalo', ['dec']],  
" 'dosadna', 'dosadna', ['negativno']",  
" '.', '.', ['.'])"]
```

Slika 4.18. Rezultat označavanja sentiment reči

Na osnovu oznaka pridruženih svakoj reči broji se koliko pozitivnih i negativnih reči sadrži analizirani dokument. Kalkulacija intenziteta iskazanog sentimenta vrši se na osnovu informacije

o prethodnom tokenu kako bi se donela odluka o intenzitetu sentimenta trenutnog tokena. Dato pravilo je posebno bitno u situacijama kada se u analiziranom dokumentu pojave negacije i druge reči koje menjaju polaritet sentimenta naredne reči, intenzifikatori i neutralizatori sentimenta. Reči koje menjaju sentiment naredne reči imaju oznaku *inv* i dodeljuje im se snažniji intenzitet sentimenta kako ne bi došlo do neutralizovanja njihove pojave u sadržaju u situacijama kada se u tekstu pojave sentiment reči suprotne semantike. Takođe, ove reči menjaju sentiment polaritet naredne reči u rečenici dodeljujući negativnu ocenu ukoliko je pozitivna sentiment reč naredna u nizu, odnosno dodeljujući pozitivnu ocenu sentimenta ukoliko im sledi negativna sentiment reč. Ukoliko se u analiziranom sadržaju pojave reči koje intenziviraju sentiment naredne reči u rečenici (u oznaci *inc*), sentiment ocena se duplira bez obzira na polaritet naredne reči, dok se pojavom reči koje umanjuju sentiment (u oznaci *dec*) sentiment ocena se deli napola. U računanju sentiment skora primenjeni su *Algoritam K.3.* i *Algoritam K.4.* prikazani u [Prilogu K.](#)

Slike 4.19. i 4.20. ilustruju rezultate analiza. Sistem prvo prikazuje recenziju ili rečenicu koja je predmet analize, zatim listu reči u rečenici sa pripadajućim oznakama tipa reči. Označene su odgovarajućim oznakama samo reči koje se nalaze u nekom od rečnika. Prva reč u nizu predstavlja originalnu vrednost reči, druga reč u nizu bi trebalo da predstavlja lemu reči, odnosno generalizovani oblik reči (npr. reči *idu*, *idem*, *idemo* bi imale generalizovan oblik *ići*), dok je poslednja vrednost u nizu oznaka na osnovu koje se vidi kojem rečniku reč pripada i na osnovu koje se vrši računanje sentiment skora, koji je prikazan u nastavku. U poslednjem delu sistem daje sentiment polaritet recenzije ili rečenice. Sistem za lematizaciju reči predstavljaće predmet budućih istraživanja jer, prema našim saznanjima, ne postoji javno dostupan sistem za srpski jezik. Očekivanja su da bi se svođenjem reči na njihov osnovni oblik postigli bolji rezultati analiza i potencijalno bi se omogućila generalizacija sentiment rečnika i na druge domene. U primeru prikazanom na slici 4.19. neutralizator *pomalo* umanjuje sentiment skor recenzije, dok u primeru prikazanom na slici 4.20. intenzifikator *veoma* uvećava intenzitet iskazanog sentimenta, a *ali* menja sentiment naredne reči u nizu.

```
Analizirana recenzija studenta:
Pomalo dosadna.

Lista svih reci u recenziji
["['pomalo', 'pomalo', ['dec']],
 " 'dosadna', 'dosadna', ['negativno']],
 " '.,', '.,', ['.']]"]

('Sentiment skor je:', -0.5)

Rezultat analize
['Student komentariše rad profesora u negativnom svetlu.']
```

Slika 4.19. Rezultat analize – primer 1

```

Analizirana recenzija studenta:
Veoma dosadna predavanja, zena zna matematiku ali samo za sebe...

Lista svih reci u recenziji
["['veoma', 'veoma', ['inc']],
 " 'dosadna', 'dosadna', ['negativno']],
 " 'predavanja', 'predavanja'",
 " ',', ',', ['']",
 " 'zena', 'zena'",
 " 'zna', 'zna', ['pozitivno']],
 " 'matematiku', 'matematiku'",
 " 'ali', 'ali', ['inv']],
 " 'samo', 'samo', ['negativno', 'dec']],
 " 'za sebe', 'za sebe', ['negativno']],
 " '...', '...', [':'])]"

('Sentiment skor je:', -2.5)

Rezultat analize
['Student komentariše rad profesora u negativnom svetlu.']

```

Slika 4.20. Rezultat analize – primer 2

Evaluacija tačnosti rada izgrađenog modela, kao i interpretacija rezultata, prikazane su u [Poglavlju 4.3.2.](#)

#### 4.1.6. Interpretacija i prikaz rezultata

Razvijeni modeli prikazani u prethodnom poglavlju primenjeni su nad anotiranim korpusom. Obučavanje modela je vršeno nad podacima iz skupa za obučavanje, dok je evaluacija njihovog rada vršena nad podacima iz skupa za testiranje, koji do ove tačke nisu bili prezentovani modelima. Raspodela korpusa na skup za obučavanje i skup za testiranje prikazana je u [Poglavlju 4.1.3.1.](#) na slici 4.9.

Načinjene anotacije atributa *Sentiment polaritet* poslužile su kao osnova za evaluaciju tačnosti klasifikacije sadržaja prema polaritetu iskazanog mišljenja. Anotacije ovog atributa nisu rađene na nivou celokupne recenzije, već na nivou rečenice ili dela rečenice u kojem je iskazan sentiment o jednom aspektu. Iz tog razloga za potrebe evaluacije rada izgrađenih modela na nivou recenzije, recenzije su grupisane prema sledećim kriterijumima:

1. Čisto pozitivne recenzije – svako iskazano mišljenje u recenziji je pozitivne orijentacije, te su i sve anotacije atributa *Sentiment polaritet* pozitivne.
2. Čisto negativne recenzije – svako iskazano mišljenje u recenziji je negativne orijentacije, te su i sve anotacije atributa *Sentiment polaritet* negativne.



3. Pretežno pozitivne recenzije – u recenziji preovlađuje pozitivan sentiment, a kriterijum klasifikovanja recenzije u ovu kategoriju je: (broj anotacija *Sentiment polaritet* = pozitivan) > (broj anotacija *Sentiment polaritet* = negativan).
4. Pretežno negativne recenzije – u recenziji preovlađuje negativan sentiment, a kriterijum klasifikovanja recenzije u ovu kategoriju je: (broj anotacija *Sentiment polaritet* = negativan) > (broj anotacija *Sentiment polaritet* = pozitivan).
5. Jednak broj anotacija – u recenziji nema preovlađujućeg sentimenta ili je podjenak broj pozitivnih i negativnih mišljenja po različitim aspektima, a kriterijum klasifikovanja recenzije u ovu kategoriju je: (broj anotacija *Sentiment polaritet* = pozitivan) = (broj anotacija *Sentiment polaritet* = negativan).

Broj recenzija identifikovanih u korpusu prema gore navedenim kategorijama prikazan je u tabeli 4.25. U daljem tekstu čisto pozitivni i pretežno pozitivni komentari su označeni kao *pozitivna klasa*, čisto negativni i pretežno negativni komentari kao *negativna klasa*. Razvijeni modeli, koji vrše klasifikaciju primenom algoritama nadgledanog učenja na nivou recenzije, koriste ove dve klase sentiment polariteta. U sentiment analizi zasnovanoj na leksikonima, u klasifikaciji na nivou dokumenta, upotrebljena je i poslednja kategorija recenzija koja je označena kao *neutralna klasa*.

Tabela 4.25. Kategorije recenzija

	Skup za obučavanje	Skup za testiranje
Čisto pozitivne	1570	610
Čisto negativne	731	243
Pretežno pozitivne	161	65
Pretežno negativne	115	58
Jednak broj	225	87

Naredna tabela prikazuje distribuciju anotiranih rečenica prema polaritetu. U klasifikaciji sentimenta na nivou rečenice, primenom algoritama mašinskog učenja, koriste se pozitivno i negativno anotirane rečenice. Razvijeni modeli sentiment analize zasnovani na leksikonima, koji vrše klasifikaciju na nivou rečenice, koriste anotirane rečenice prema sve tri klase sentiment polariteta.

Tabela 4.26. Broj anotiranih rečenica prema polaritetu

	Skup za obučavanje	Skup za testiranje
Pozitivne	3490	1451
Negativne	2117	934
Neutralne	87	22

Upotrebene mere tačnosti su:

1. **Acc** – *Accuracy* predstavlja tačnost klasifikacije:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}$$

2. **P** – *Precision* ili preciznost klasifikacije predstavlja procenat recenzija koje je model tačno klasifikovao, odnosno udeo stvarno pozitivnih primera u svim primerima koji su modelom predviđeni kao pozitivni.

$$P = \frac{TP}{TP + FP}$$

3. **R** – *Recall* ili preklapanje. Od ukupnog broja recenzija koje stvarno pripadaju određenoj klasi, *recall* označava procenat recenzija koje je sistem tačno klasifikovao.

$$R = \frac{TP}{TP + FN}$$

Zavisno od toga da li je za konkretnu aplikaciju bitnije da pronađe sve recenzije koje pripadaju određenoj klasi ili celokupna tačnost klasifikacije, vrednosti mera *P* i *R* se mogu varirati kako bi se povećao značaj jedne ili druge mere. F-mera je uvedena kako bi se kombinovale prethodne dve mere i omogućila uporedivost različitih sistema.

4. **F<sub>1</sub> - mera** – predstavlja harmonijsku sredinu preciznosti *P* i preklapanja *R*.

$$F_1 - \text{mera} = \frac{2 \times P \times R}{P + R}$$

Argumenti navedenih mera se generišu na osnovu matrice koja se kreira za svaku analiziranu klasu u sledećoj formi:

Tabela 4.27. Struktura matrice konfuzije

	Stvarne vrednosti	
	Tačno	nije tačno
Sistemski odabrano	TP	FP
Sistem nije odabrao	FN	TN

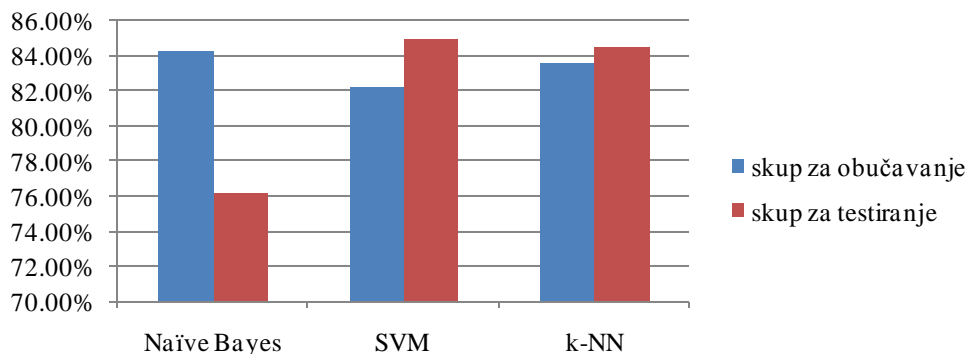
gde  $TP$  predstavlja stvarne pozitivne vrednosti,  $FN$  lažne negativne vrednosti,  $FP$  lažne pozitivne vrednosti i  $TN$  stvarne negativne vrednosti.

#### 4.1.6.1. Evaluacija klasifikatora izgrađenih algoritmima mašinskog učenja

Klasifikacija sentimenta primenom nadgledanih tehnika mašinskog učenja koristi dva skupa anotiranih podataka za obučavanje klasifikatora: klasu pozitivnih komentara, odnosno rečenica, i klasu negativnih komentara, odnosno rečenica. Klasifikacija je vršena na nivou celog dokumenta, odnosno studentske recenzije i na nivou pojedinačnih rečenica. Evaluacija razvijenih modela izvršena je nad podacima iz skupa za testiranje. U nastavku poglavlja ilustrovana je preciznost rada modela.

##### Klasifikacija na nivou dokumenta

Slika 4.21. prikazuje sveukupnu tačnost klasifikacije na nivou dokumenta, odnosno celokupne recenzije, iskazanu merom *Accuracy*. Može se zapaziti da nad skupom za obučavanje Naïve Bayes algoritam ima najbolje performanse: postiže tačnost klasifikacije od 84.31%. Međutim, nad test podacima postiže značajno lošiju tačnost od 76.13%. Najvišu tačnost klasifikacije nad test podacima postiže SVM klasifikator od 84.94% tačnosti. Sveukupno posmatrano, algoritam k-najbližih suseda daje podjednako dobre rezultate nad oba skupa. Tabele u nastavku poglavlja ilustruju preciznost rada razvijenih modela nad celokupnim recenzijama studenata iskazane merama *Precision*, *Recall* i *F-merom*. Na ovaj način stiže se bolji uvid o klasifikaciji pojedinih modela po klasama sentimenta i omogućuje se uporedivost modela.



Slika 4.21. Tačnost rada modela – mera Accuracy

Klasifikacija Naive Bayes klasifikatorom na nivou dokumenta

Tabela 4.28. ilustruje preciznost rada klasifikatora nad podacima za obučavanje. Na osnovu prezentovanih podataka može se zaključiti da se klasifikacijom recenzija Naive Bayes klasifikatorom postiže skladan odnos mera *Precision* i *Recall* i da se uspešno prepoznaju pozitivni i negativni komentari. Interesantno je primetiti da model ispoljava značajno lošije performanse nad test podacima (tabela 4.29). Preciznost klasifikacije negativnih komentara pada sa 74.64% na 58.25%, dok se *Recall* mera ne menja značajno. Po pitanju pozitivne klase, vrednost mere *Recall* pada za više od 10%. Sveukupno posmatrano, ukoliko se modelu prezentuju novi, njemu nepoznati podaci, pozitivni komentari će biti prepoznati u 81.19% slučajeva, dok će negativni komentari biti prepoznati u svega 67.32% slučajeva (*F-mera* prikazana u poslednjem redu obe tabele).

Tabela 4.28. Preciznost rada Naive Bayes klasifikatora nad podacima za obučavanje

Accuracy: 84.31% +/- 1.75% (mikro: 84.31%)			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1500	176	89.50%
Predviđeno negativno	228	671	74.64%
Recall	86.81%	79.22%	
<b>F-mera</b>	<b>88.13%</b>	<b>76.86%</b>	

Tabela 4.29. Preciznost rada Naive Bayes klasifikatora nad podacima za testiranje

Accuracy: 76.13%			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	503	61	89.18%
Predviđeno negativno	172	240	58.25%
Recall	74.52%	79.73%	
<b>F-mera</b>	<b>81.19%</b>	<b>67.32%</b>	

Klasifikacija Support vector machines klasifikatorom na nivou dokumenta

Preciznost rada SVM klasifikatora nad podacima za obučavanje i nad podacima za testiranje je prikazana u tabelama 4.30. i 4.31. respektivno. Za razliku od Naive Bayes klasifikatora SVM klasifikator postiže sveukupno precizniju klasifikaciju obe klase. U 89.66% slučajeva klasifikator će uspešno klasifikovati pozitivnu klasu, dok će u 72.32% slučajeva uspešno klasifikovati negativnu klasu; podaci su prikazani u poslednjem redu tabele 4.31. Takođe, preciznost identifikovanja negativne klase Naive Bayes klasifikatorom (58.25%) značajno je niža u poređenju sa SVM klasifikatorom (83.48%), dok je *Recall* niži u klasifikaciji sa SVM klasifikatorom za blizu 16%.

**Tabela 4.30. Preciznost rada SVM klasifikatora nad podacima za obučavanje**

<b>Accuracy: 82.25% +/- 3.05% (mikro: 82.25%)</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1606	335	82.74%
Predviđeno negativno	122	512	80.76%
Recall	92.94%	60.45%	
<b>F-mera</b>	<b>87.54%</b>	<b>69.14%</b>	

**Tabela 4.31. Preciznost rada SVM klasifikatora nad podacima za testiranje**

<b>Accuracy: 84.94%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	637	109	85.39%
Predviđeno negativno	38	192	83.48%
Recall	94.37%	63.79%	
<b>F-mera</b>	<b>89.66%</b>	<b>72.32%</b>	

Klasifikacija k-Nearest Neighbor klasifikatorom na nivou dokumenta

Preciznost rada k-NN klasifikatora nad podacima za obučavanje i nad podacima za testiranje je prikazana u tabelama 4.32. i 4.33. respektivno.

**Tabela 4.32. Preciznost rada k-NN klasifikatora nad podacima za obučavanje**

<b>Accuracy: 83.65% +/- 1.58% (mikro: 83.65%)</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1597	290	84.63%
Predviđeno negativno	131	557	80.96%
Recall	92.42%	65.76%	
<b>F-mera</b>	<b>88.35%</b>	<b>72.57%</b>	

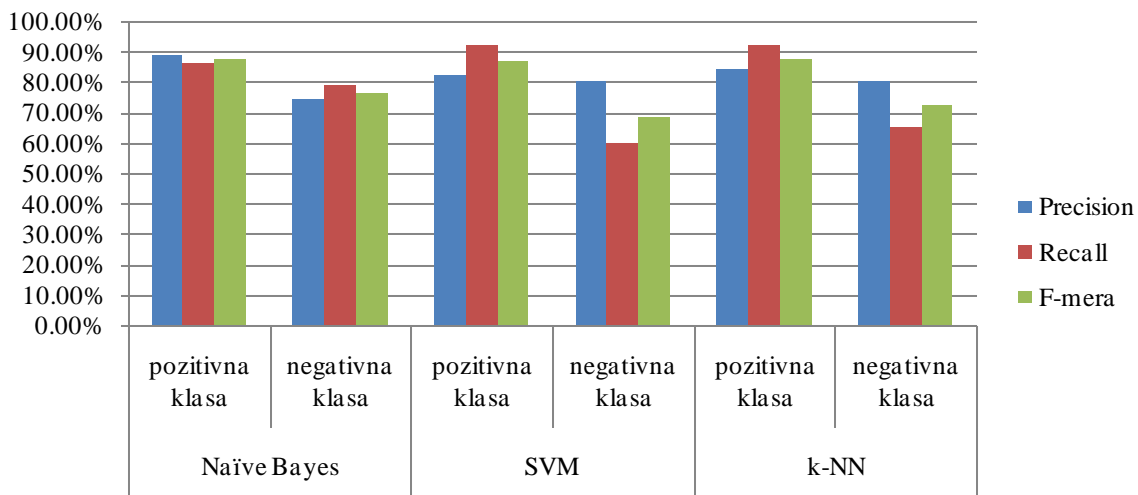
**Tabela 4.33. Preciznost rada k-NN klasifikatora nad podacima za testiranje**

<b>Accuracy: 84.53%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	607	83	87.97%
Predviđeno negativno	68	218	76.22%
Recall	89.93%	72.43%	
<b>F-mera</b>	<b>88.94%</b>	<b>74.28%</b>	

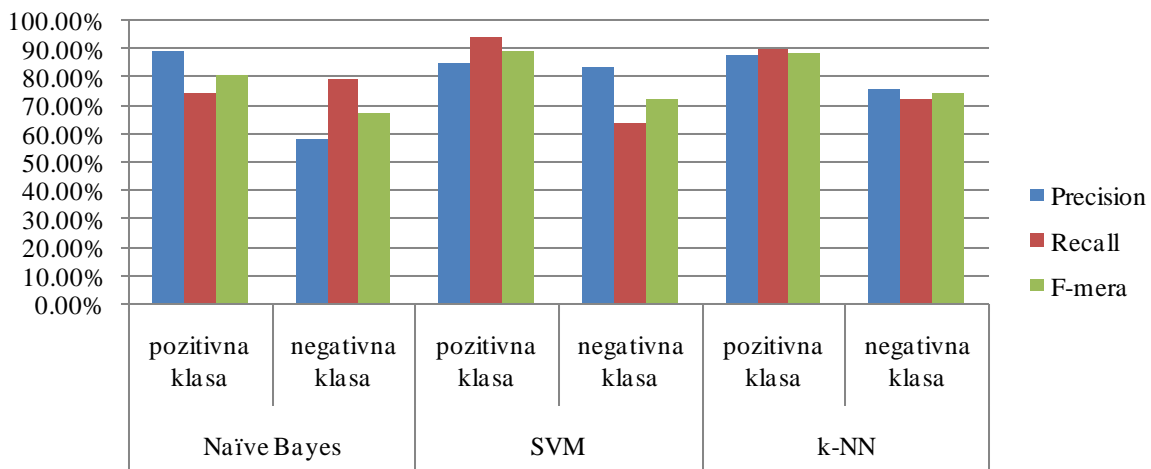
U poređenju sa prethodna dva klasifikatora, k-NN klasifikator daje najbolji kompromis mera *Precision* i *Recall* nad test podacima. Prema podacima prikazanim u tabeli 4.33. vidi se da je

preciznost identifikovanja pozitivnih komentara 87.97%, a *Recall* ili preklapanje 89.93%. Preciznost identifikovanja negativnih komentara je 76.22%, a *Recall* 72.43%. Sveukupna tačnost klasifikacije iskazana *F-merom* pokazuje da u 88.94% slučajeva k-NN klasifikator tačno klasifikuje pozitivne komentare, dok u 74.28% slučajeva tačno klasifikuje negativne komentare.

Naredne dve slike daju paralelni prikaz mera *Precision*, *Recall* i *F-mera* za tri razvijena modela klasifikacije celokupnih recenzija studenata. Slika 4.22. prikazuje preciznost njihovog rada nad skupom za obučavanje, dok slika 4.23. prikazuje preciznost njihovog rada nad skupom za testiranje. Sva tri razvijena modela sa manje uspeha identifikuju komentare negativne sentiment orijentacije u odnosu na pozitivne.



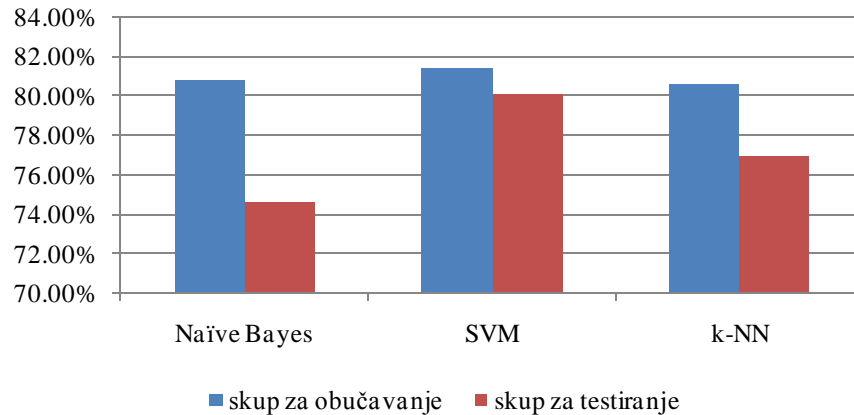
Slika 4.22. Preciznost rada modela nad podacima za obučavanje



Slika 4.23. Preciznost rada modela nad podacima za testiranje

**Klasifikacija na nivou rečenice**

Na nivou rečenice najviša tačnost klasifikacije postiže se SVM klasifikatorom. Na slici 4.24. prikazana je tačnost klasifikacije tri modela iskazana merom *Accuracy* nad podacima za obučavanje i testiranje.



**Slika 4.24. Tačnost rada modela – mera Accuracy**

Naredne tabele ilustruju preciznost klasifikacije na nivou rečenica, iskazanu merama *Precision*, *Recall* i *F-merom*. Na ovaj način stiže se bolji pregled klasifikacije pojedinih modela po klasama sentimenta i omogućuje se uporedivost modela.

**Klasifikacija Naïve Bayes klasifikatorom na nivou rečenice**

Naïve Bayes klasifikator ispoljava nižu tačnost klasifikacije negativnih rečenica iz test skupa (63.12%) u poređenju sa preciznošću klasifikacije rečenica iz skupa za obučavanje (72.02%). Na osnovu podataka prikazanih u tabelama 4.34. i 4.35. može se takođe primetiti drastičan pad vrednosti *Recall* mere za pozitivnu klasu – sa 80.97% u klasifikaciji rečenica iz skupa za obučavanje na 68.02% u klasifikaciji rečenica iz skupa za testiranje. Sveukupno posmatrano, Naïve Bayes klasifikator će identifikovati pozitivne rečenice u 76.57% slučajeva, dok će negativne rečenice prepoznati u 72.45% slučajeva.

**Tabela 4.34. Preciznost rada Naïve Bayes klasifikatora nad podacima za obučavanje**

Accuracy: 80.88% +/- 2.10% (mikro: 80.88%)			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	2826	408	87.38%
Predviđeno negativno	664	1709	72.02%
Recall	80.97%	80.73%	
<b>F-mera</b>	<b>84.05%</b>	<b>76.13%</b>	

**Tabela 4.35. Preciznost rada Naïve Bayes klasifikatora nad podacima za testiranje**

Accuracy: 74.68%			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	987	140	87.58%
Predviđeno negativno	464	794	63.12%
Recall	68.02%	85.01%	
<b>F-mera</b>	<b>76.57%</b>	<b>72.45%</b>	

Klasifikacija Support vector machines klasifikatorom na nivou rečenice

Na osnovu podataka prikazanih u tabelama 4.36. i 4.37. može se zaključiti da SVM klasifikator ima približne performanse nad skupom za obučavanje i nad skupom za testiranje. Sveukupno posmatrano, SVM klasifikator će uspešno prepoznati pozitivne rečenice u 84.53% slučajeva, a negativne rečenice u 72.21% slučajeva.

**Tabela 4.36. Preciznost rada SVM klasifikatora nad podacima za obučavanje**

Accuracy: 81.50% +/- 1.03% (mikro: 81.51%)			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	3145	692	81.97%
Predviđeno negativno	345	1425	80.51%
Recall	90.11%	67.31%	
<b>F-mera</b>	<b>85.85%</b>	<b>73.32%</b>	

**Tabela 4.37. Preciznost rada SVM klasifikatora nad podacima za testiranje**

Accuracy: 80.13%			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1295	318	80.29%
Predviđeno negativno	156	616	79.79%
Recall	89.25%	65.95%	
<b>F-mera</b>	<b>84.53%</b>	<b>72.21%</b>	

Klasifikacija k-Nearest Neighbor klasifikatorom na nivou rečenice

Preciznost rada k-NN klasifikatora nad podacima za obučavanje i nad podacima za testiranje je prikazana u tabelama 4.38. i 4.39. respektivno. k-NN klasifikator pozitivnu klasu tačno identifikuje u 81.50% slučajeva, dok negativnu uspešno identifikuje u 69.69% slučajeva, kao što ilustruju podaci prikazani u tabeli 4.39. Takođe se može zapaziti da ne postoje značajnije oscilacije u radu klasifikatora nad podacima za obučavanje i nad podacima za testiranje.



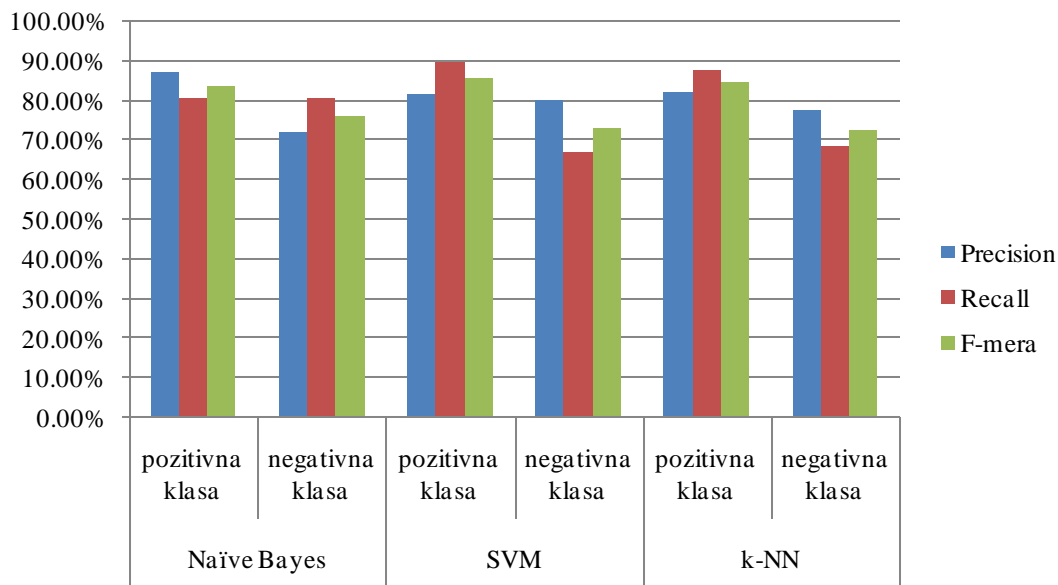
**Tabela 4.38. Preciznost rada k-NN klasifikatora nad podacima za obučavanje**

Accuracy: 80.67% +/- 1.25% (mikro: 80.67%)			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	3069	663	82.23%
Predviđeno negativno	421	1454	77.55%
Recall	87.94%	68.68%	
<b>F-mera</b>	<b>84.99%</b>	<b>72.85%</b>	

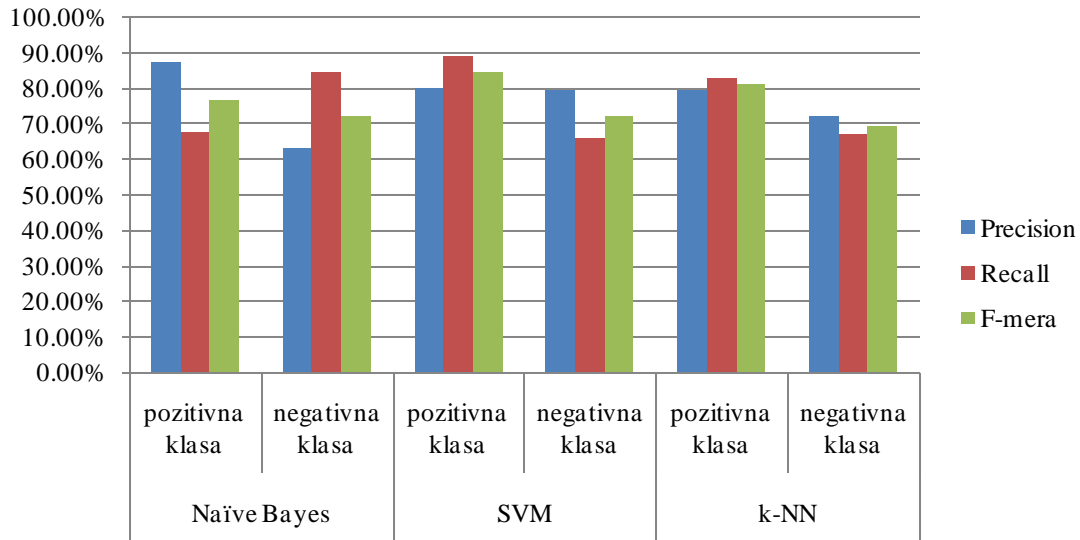
**Tabela 4.39. Preciznost rada k-NN klasifikatora nad podacima za testiranje**

Accuracy: 77.02%			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1207	304	79.88%
Predviđeno negativno	244	630	72.08%
Recall	83.18%	67.45%	
<b>F-mera</b>	<b>81.50%</b>	<b>69.69%</b>	

Naredne dve slike daju paralelni prikaz mera *Precision*, *Recall* i *F-mere* za tri razvijena modela klasifikacije na nivou rečenica. Slika 4.25. prikazuje preciznost njihovog rada nad skupom za obučavanje, dok slika 4.26. prikazuje preciznost njihovog rada nad skupom za testiranje. Kao i u slučaju klasifikacije na nivou dokumenta, može se zapaziti da sva tri razvijena modela manje uspešno identifikuju rečenice negativne sentiment orijentacije u odnosu na pozitivne.



**Slika 4.25. Preciznost rada modela nad podacima za obučavanje**



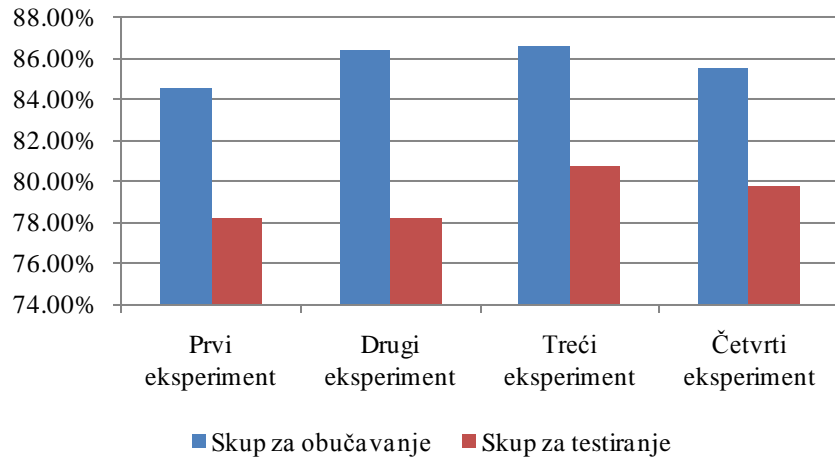
Slika 4.26. Preciznost rada modela nad podacima za testiranje

#### 4.1.6.2. Evaluacija modela sentiment analize zasnovane na leksikonima

Razvijeni model sentiment analize baziran na rečnicima i lingvističkim pravilnostima primenjen je dvostruko, nad celokupnim recenzijama i nad pojedinačnim rečenicama. Koristi tri skupa anotiranih podataka za obučavanje klasifikatora: klasu pozitivnih komentara, odnosno rečenica, klasu negativnih komentara, odnosno rečenica, i klasu neutralnih komentara, odnosno rečenica. Evaluacija razvijenih modela izvršena je nad podacima iz skupa za testiranje, koji do ove tačke nisu prezentovani modelima. U nastavku poglavlja ilustrovana je preciznost rada modela.

#### Primena modela nad recenzijama

Na nivou celokupnih recenzija sprovedena su četiri eksperimenta. U prvom eksperimentu su upotrebljeni rečnici pozitivnih i negativnih sentiment reči za klasifikaciju recenzija. U drugom eksperimentu su iz rečnika pozitivnih i negativnih sentiment reči manuelno uklanjani pojmovi koje su anotatori označili kao sentiment reč, a koje ne iskazuju sentiment. U narednom eksperimentu su za klasifikaciju recenzija upotrebljeni manuelno korigovani rečnici pozitivnih i negativnih sentiment reči, ali i rečnik negacije i drugih reči koje menjaju sentiment polaritet naredne reči, kao i rečnici pojmova koji pojačavaju i umanjuju iskazani sentiment. U četvrtom eksperimentu su korišćeni isti rečnici kao u trećem eksperimentu, pri čemu je ojačan sentiment pojmova koji menjaju sentiment polaritet naredne reči. Slika 4.27. ilustruje tačnost klasifikacije modela u sva četiri eksperimenta, iskazanu merom *Accuracy*. Može se zaključiti da treći eksperiment daje najbolje rezultate. U nastavku su detaljnije opisani realizovani eksperimenti.



Slika 4.27. Tačnost modela - mera Accuracy

*Prvi eksperiment:* Upotrebljeni su rečnici pozitivnih i negativnih sentiment reči pre manuelne korekcije i uklanjanja pojmova koje su anotatori označili kao sentiment reč, a koje ne iskazuju sentiment, poput *kurs*, *predavati*, *konsultacija*, *komunikacija*. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje. Na osnovu podataka prikazanih u tabelama 4.40. i 4.41. može se zapaziti da dolazi do značajnog pada vrednosti mere *Recall* za negativnu klasu, sa 85.58% nad skupom za obučavanje na 69.77% nad skupom za testiranje. Takođe, preciznost i tačnost identifikovanja komentara pozitivne semantike izuzetno je visoka.

Tabela 4.40. Eksperiment 1: Preciznost rada modela – skup za obučavanje

Accuracy: 84.55%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1581	40	84	92.73%
Predviđeno negativno	50	724	77	85.08%
Predviđeno neutralno	100	82	64	26.02%
Recall	91.33%	85.58%	28.44%	
<b>F-mera</b>	<b>92.03%</b>	<b>85.34%</b>	<b>27.18%</b>	

Tabela 4.41. Eksperiment 1: Preciznost rada modela – skup za testiranje

Accuracy: 78.25%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	597	39	37	88.71%
Predviđeno negativno	31	210	26	78.65%
Predviđeno neutralno	46	52	24	19.67%
Recall	88.58%	69.77%	27.59%	
<b>F-mera</b>	<b>88.64%</b>	<b>73.94%</b>	<b>22.97%</b>	

*Drugi eksperiment:* Upotrebljeni su rečnici pozitivnih i negativnih sentiment reči nakon manuelne korekcije, koja je preciznije opisana u [Poglavlju 4.1.4](#). Na osnovu podataka prikazanih u tabelama 4.42. i 4.43. može se primetiti porast tačnosti klasifikacije kada je u pitanju skup za obučavanje, međutim, preciznost rada modela nad skupom za testiranje ostaje nepromenjena.

**Tabela 4.42. Eksperiment 2: Preciznost rada modela – skup za obučavanje**

Accuracy: 86.40%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1647	49	98	91.81%
Predviđeno negativno	30	704	57	89%
Predviđeno neutralno	54	93	70	32.26%
Recall	95.15%	83.22%	31.11%	
<b>F-mera</b>	<b>93.45%</b>	<b>86.01%</b>	<b>31.67%</b>	

**Tabela 4.43. Eksperiment 2: Preciznost rada modela – skup za testiranje**

Accuracy: 78.25%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	610	46	43	87.27%
Predviđeno negativno	20	191	21	82.33%
Predviđeno neutralno	45	54	23	18.85%
Recall	90.37%	65.64%	26.44%	
<b>F-mera</b>	<b>88.79%</b>	<b>73.04%</b>	<b>22.01%</b>	

*Treći eksperiment:* Upotrebljeni su manuelno korigovani rečnici pozitivnih i negativnih sentiment reči, rečnik negacije i drugih reči koje menjaju sentiment polaritet naredne reči, kao i rečnici pojmova koji pojačavaju i umanjuju iskazani sentiment. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje. Uvođenje novih rečnika dovodi do poboljšanja preciznosti i tačnosti identifikovanja novih, nepoznatih primera, kako pozitivnih tako i negativnih. Negativna klasa je posebno značajna jer ona oslikava nezadovoljstvo studenata i ukazuje na aspekte u kojima bi trebalo raditi na poboljšanju. Otuda je značajno povećanje *F-mere* sa 73.04% u prethodnom eksperimentu na 77.28%.

**Tabela 4.44. Eksperiment 3: Preciznost rada modela – skup za obučavanje**

Accuracy: 86.62%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1664	55	108	91.08%
Predviđeno negativno	24	709	63	89.07%
Predviđeno neutralno	43	82	54	30.17%
Recall	96.13%	83.81%	24%	
<b>F-mera</b>	<b>93.54%</b>	<b>86.36%</b>	<b>26.73%</b>	

Tabela 4.45. Eksperiment 3: Preciznost rada modela – skup za testiranje

Accuracy: 80.71%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	617	40	43	88.14%
Predviđeno negativno	23	216	19	83.72%
Predviđeno neutralno	35	45	25	23.81%
Recall	91.41%	71.76%	28.74%	
<b>F-mera</b>	<b>89.75%</b>	<b>77.28%</b>	<b>26.04%</b>	

*Četvrti eksperiment:* Upotrebljeni su manuelno korigovani rečnici pozitivnih i negativnih sentiment reči, rečnici pojmova koji pojačavaju i umanjuju iskazani sentiment, kao i rečnik pojmova koji menjaju sentiment polaritet naredne reči, pri čemu je ojačan sentiment ovih pojmova prema Algoritmu K.3, prikazanom u *Prilogu K* i opisanom u prethodnom poglavlju. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje. Može se zaključiti da navedene izmene ne donose poboljšanje performansi modela.

Tabela 4.46. Eksperiment 4: Preciznost rada modela – skup za obučavanje

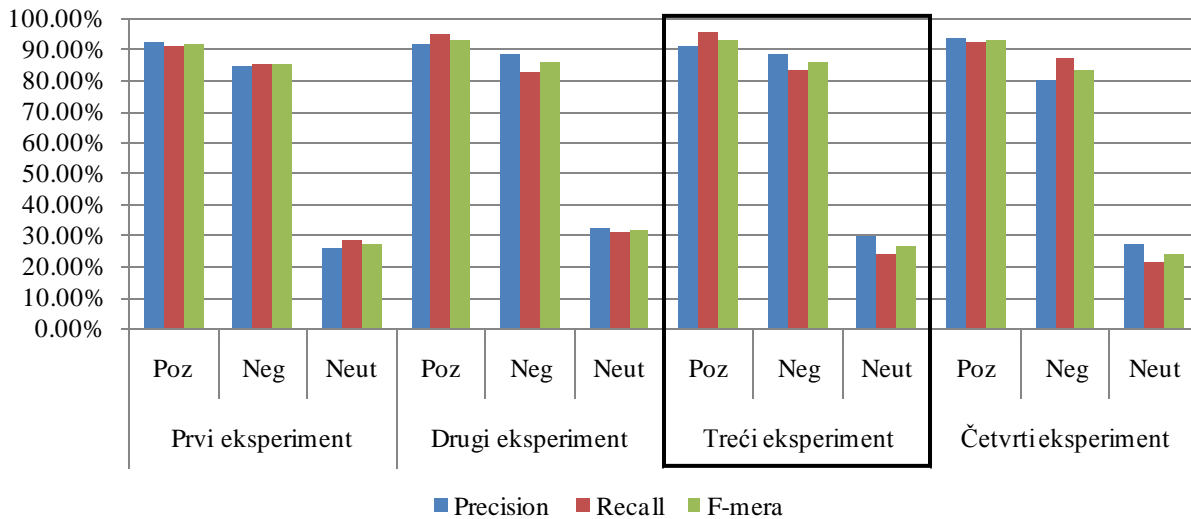
Accuracy: 85.51%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1609	41	61	94.04%
Predviđeno negativno	62	739	116	80.59%
Predviđeno neutralno	60	66	48	27.59%
Recall	92.95%	87.35%	21.33%	
<b>F-mera</b>	<b>93.49%</b>	<b>83.83%</b>	<b>24.06%</b>	

Tabela 4.47. Eksperiment 4: Preciznost rada modela – skup za testiranje

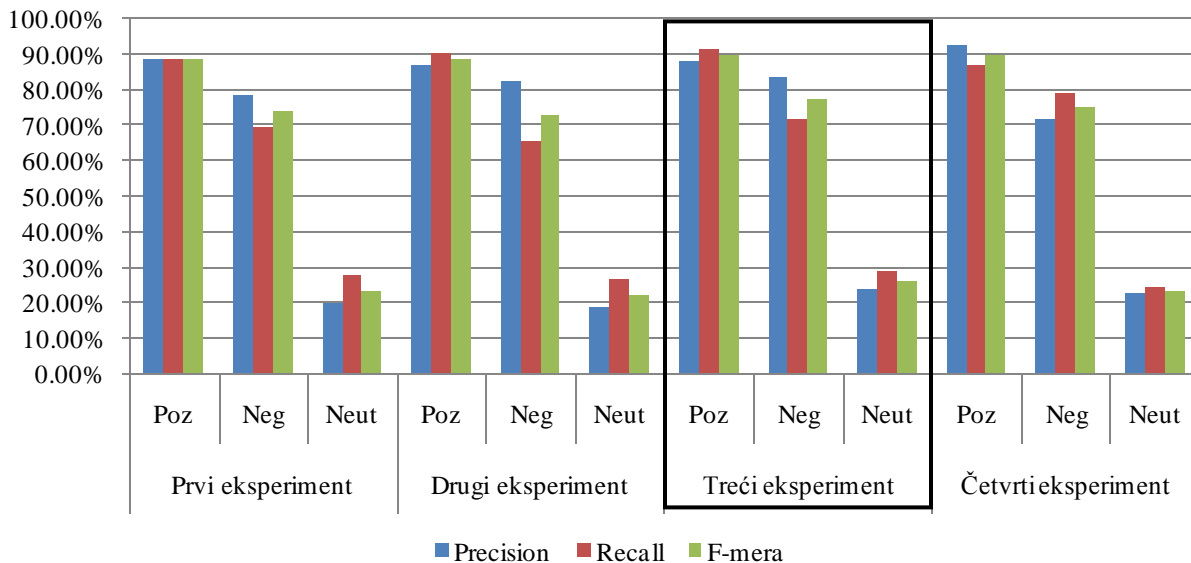
Accuracy: 79.77%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	588	25	23	92.45%
Predviđeno negativno	52	239	43	71.56%
Predviđeno neutralno	35	37	21	22.58%
Recall	87.11%	79.40%	24.14%	
<b>F-mera</b>	<b>89.70%</b>	<b>75.28%</b>	<b>23.33%</b>	

Naredne dve slike sumiraju preciznost rada izgrađenog modela nad podacima iz skupa za obučavanje i nad podacima iz skupa za testiranje, paralelno za sve realizovane eksperimente. Na osnovu podataka može se zaključiti da se preciznost rada izgrađenog modela povećava uvođenjem dodatnih rečnika (modifikatora i reči koje menjaju sentiment polaritet naredne reči

u tekstu). Promene su neznatne na nivou recenzija iz skupa za obučavanje, međutim, značajne kada je reč o test podacima, a prevashodno kada je reč o negativnoj klasi. Takođe, na osnovu ilustracija može se приметiti da sa povećanjem mere *Recall* dolazi do pada mere *Precision* i obrnuto.



Slika 4.28. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za obučavanje



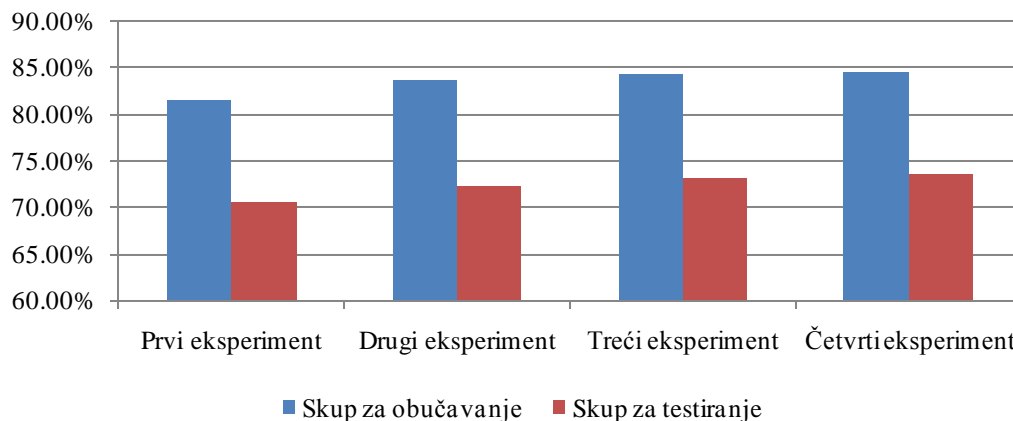
Slika 4.29. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za testiranje

Treći eksperiment daje najbolje rezultate nad test podacima, kao što prikazuje slika 4.29. Izgrađeni model će u 89.75% slučajeva identifikovati pozitivan komentar, u 77.28% slučajeva negativan, dok u svega 26.04% neutralan. U analizi sentimenta u komentarima studenata,

negativan sentiment predstavlja posebno značajnu klasu jer se u negativnim komentarima može identifikovati nezadovoljstvo studenata. Detaljnom analizom ovih sadržaja identifikovaće se uska grla i problemi sa kojima se studenti susreću, te će se moći raditi na unapređenju poslovanja. Otuda je za primenu razvijenog modela donekle bitnija mera *Precision*, odnosno procenat slučajeva u kojima će model tačno identifikovati određenu klasu. Ukoliko se ova mera posmatra samostalno, najbolji rezultati po pitanju negativne klase postignuti su takođe u trećem eksperimentu: u 83.72% slučajeva model će tačno identifikovati negativnu recenziju.

### Primena modela nad rečenicama

Razvijeni model sentiment analize zasnovane na leksikonima primenjen je nad anotiranim rečenicama u četiri eksperimenta koji kombinuju različite rečnike. U prvom eksperimentu su za klasifikaciju rečenica upotrebljeni manuelno korigovani rečnici pozitivnih i negativnih sentiment reči. U drugom eksperimentu pored sentiment rečnika upotrebljen je rečnik negacije i drugih reči koje menjaju sentiment polaritet naredne reči. U narednom eksperimentu su upotrebljeni manuelno korigovani rečnici pozitivnih i negativnih sentiment reči, ali i rečnik negacije i drugih reči koje menjaju sentiment polaritet naredne reči, kao i rečnici pojmova koji pojačavaju i umanjuju iskazani sentiment. U četvrtom eksperimentu su pri klasifikaciji rečenica korišćeni isti rečnici kao u trećem eksperimentu, ali je ojačan sentiment pojmova koji menjaju sentiment polaritet naredne reči. Na osnovu slike 4.30. može se zaključiti da treći i četvrti ekperiment daju približnu tačnost klasifikacije, iskazanu merom *Accuracy*. U nastavku su detaljnije opisani realizovani eksperimenti.



Slika 4.30. Tačnost modela - mera *Accuracy*

Prvi eksperiment: Upotrebljeni su manuelno korigovani rečnici pozitivnih i negativnih sentiment pojmova identifikovanih u fazi anotiranja korpusa. Na osnovu podataka prikazanih u tabelama

4.48. i 4.49. može se zaključiti da dolazi do drastičnog pada u vrednostima *Recall* za negativnu klasu. Vrednost mere *Recall* pada sa 76.85% u klasifikaciji rečenica iz skupa za obučavanje na 57.82% u klasifikaciji rečenica iz skupa za testiranje.

**Tabela 4.48. Eksperiment 1: Preciznost rada modela – skup za obučavanje**

Accuracy: 81.54%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	2975	116	18	95.69%
Predviđeno negativno	125	1627	28	91.40%
Predviđeno neutralno	390	374	41	5.09%
Recall	85.24%	76.85%	47.13%	
<b>F-mera</b>	<b>90.16%</b>	<b>83.50%</b>	<b>9.19%</b>	

**Tabela 4.49. Eksperiment 1: Preciznost rada modela – skup za testiranje**

Accuracy: 70.59%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1148	113	4	90.75%
Predviđeno negativno	96	540	7	83.98%
Predviđeno neutralno	207	281	11	2.20%
Recall	79.12%	57.82%	50%	
<b>F-mera</b>	<b>84.54%</b>	<b>68.49%</b>	<b>4.21%</b>	

*Drugi eksperiment:* Upotrebljeni su manuelno korigovani rečnici jedinstvenih pozitivnih i negativnih sentiment pojmova, kao i rečnik negacije i drugih reči koje menjaju sentiment polaritet naredne reči u rečenici. Uvođenje navedenih promena ne dovodi do značajnih pomaka u preciznosti rada modela.

**Tabela 4.50. Eksperiment 2: Preciznost rada modela – skup za obučavanje**

Accuracy: 83.81%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	3155	160	17	94.69%
Predviđeno negativno	73	1572	25	94.13%
Predviđeno neutralno	262	385	45	6.50%
Recall	90.40%	74.26%	51.72%	
<b>F-mera</b>	<b>92.50%</b>	<b>83.02%</b>	<b>11.55%</b>	



**Tabela 4.51. Eksperiment 2: Preciznost rada modela – skup za testiranje**

Accuracy: 72.29%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1197	119	5	90.61%
Predviđeno negativno	71	531	5	87.48%
Predviđeno neutralno	183	284	12	2.51%
Recall	82.49%	56.85%	54.55%	
<b>F-mera</b>	<b>86.36%</b>	<b>68.92%</b>	<b>4.79%</b>	

*Treći eksperiment:* Upotrebljeni su manuelno korigovani rečnici jedinstvenih pozitivnih i negativnih sentiment reči identifikovanih u fazi anotacije korpusa, rečnik negacije i drugih pojmova koji menjaju sentiment polaritet naredne reči u rečenici i rečnici pojmova koji pojačavaju, odnosno umanjuju intenzitet iskazanog sentimenta. Na osnovu podataka prikazanih u tabelama 4.52. i 4.53. može se zaključiti da uvođenje novih rečnika dovodi do pomaka u preciznosti klasifikacije pozitivnih i negativnih rečenica. U 86.50% slučajeva model uspešno klasifikuje pozitivne rečenice, dok u 70.03% slučajeva uspešno klasifikuje negativne rečenice. Podaci ukazuju na značajno smanjenje u preciznosti i tačnosti klasifikovanja neutralnih rečenica, *F-mera* pada sa 12.21% na 5.04%.

**Tabela 4.52. Eksperiment 3: Preciznost rada modela – skup za obučavanje**

Accuracy: 84.42%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	3162	160	17	94.70%
Predviđeno negativno	80	1600	25	93.84%
Predviđeno neutralno	248	357	45	6.92%
Recall	90.60%	75.58%	51.72%	
<b>F-mera</b>	<b>92.60%</b>	<b>83.73%</b>	<b>12.21%</b>	

**Tabela 4.53. Eksperiment 3: Preciznost rada modela – skup za testiranje**

Accuracy: 73.10%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1201	120	5	90.57%
Predviđeno negativno	76	548	5	87.12%
Predviđeno neutralno	174	268	12	2.64%
Recall	82.77%	58.55%	54.55%	
<b>F-mera</b>	<b>86.49%</b>	<b>70.03%</b>	<b>5.04%</b>	

*Četvrti eksperiment:* Upotrebljeni su manuelno korigovani rečnici pozitivnih i negativnih sentiment reči, rečnici pojmova koji pojačavaju i umanjuju iskazani sentiment, kao i rečnik pojmova koji menjaju sentiment polaritet naredne reči, pri čemu je ojačan sentiment ovih pojmova prema Algoritmu K.3, prikazanom u *Prilogu K*. Prezentovani model daje najvišu preciznost klasifikacije negativnih rečenica, koje predstavljaju posebnu bitnu kategoriju, *F-mera* iznosi 71.80%.

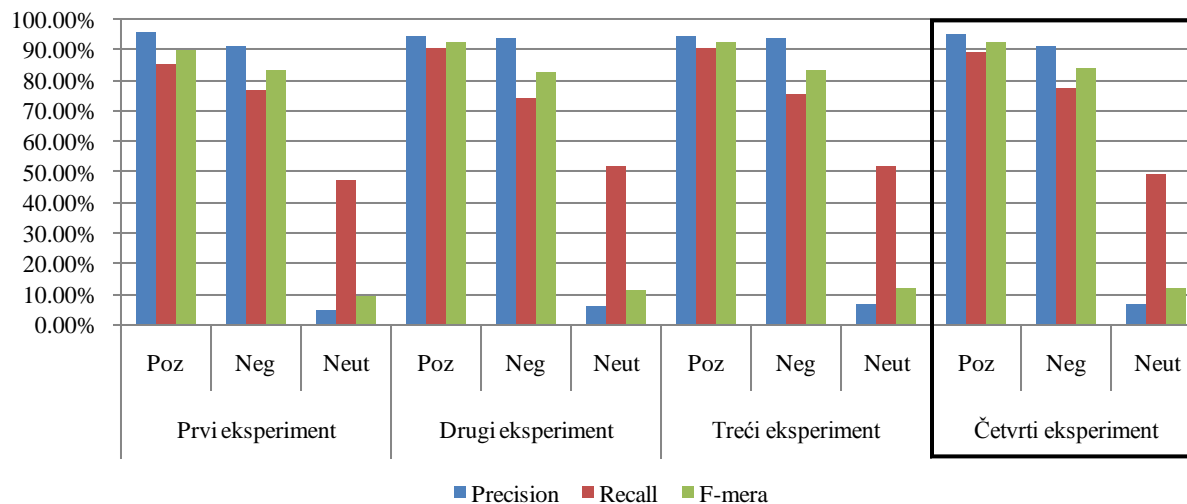
**Tabela 4.54. Eksperiment 4: Preciznost rada modela – skup za obučavanje**

Accuracy: 84.56%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	3125	140	16	95.25%
Predviđeno negativno	127	1647	28	91.40%
Predviđeno neutralno	238	330	43	7.04%
Recall	89.54%	77.80%	49.43%	
<b>F-mera</b>	<b>92.31%</b>	<b>84.05%</b>	<b>12.32%</b>	

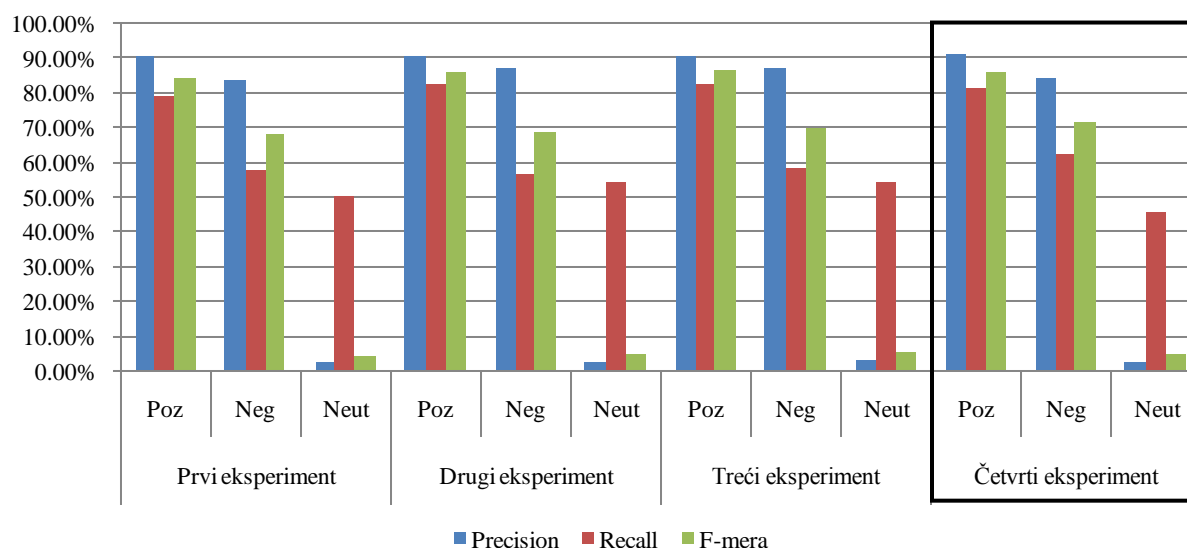
**Tabela 4.55. Eksperiment 4: Preciznost rada modela – skup za testiranje**

Accuracy: 73.70%				
	Stvarno pozitivno	Stvarno negativno	Stvarno neutralno	Precision
Predviđeno pozitivno	1181	108	5	91.27%
Predviđeno negativno	100	583	7	84.49%
Predviđeno neutralno	170	243	10	2.36%
Recall	81.39%	62.42%	45.45%	
<b>F-mera</b>	<b>86.05%</b>	<b>71.80%</b>	<b>4.49%</b>	

Naredne dve slike prikazuju preciznost rada izgrađenih modela nad podacima iz skupa za obučavanje i nad podacima iz skupa za testiranje, paralelno za sve realizovane eksperimente. Na osnovu podataka može se zaključiti da uvođenje rečnika reči koje menjaju sentiment polaritet naredne reči u rečenici uvećava tačnost klasifikacije pozitivnih i neutralnih rečenica za 2-3%, dok se tačnost klasifikacije negativnih rečenica neznatno menja. Nad skupom za testiranje *F-mera* iznosi 68.92%, međutim, preciznost klasifikatora (*Precision*) uvećava se sa 83.98%, koliko je iznosila u prvom eksperimentu, na 87.48%, što govori da će u 87.48% nepoznatih slučajeva klasifikator uspešno prepoznati negativnu rečenicu. Uvođenjem rečnika pojmova koji jačaju ili umanjuju iskazani sentiment tačnost klasifikovanja negativnih rečenica podiže se na 83.73% nad skupom za obučavanje i 70.03% nad skupom za testiranje. Četvrti eksperiment, u kojem je pojačan intenzitet reči koje menjaju polaritet, daje najbolje rezultate po pitanju identifikovanja rečenica u kojima je iskazan negativan sentiment, *F-mera* iznosi 84.05% nad skupom za obučavanje i 71.80% nad skupom za testiranje.



Slika 4.31. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za obučavanje



Slika 4.32. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za testiranje

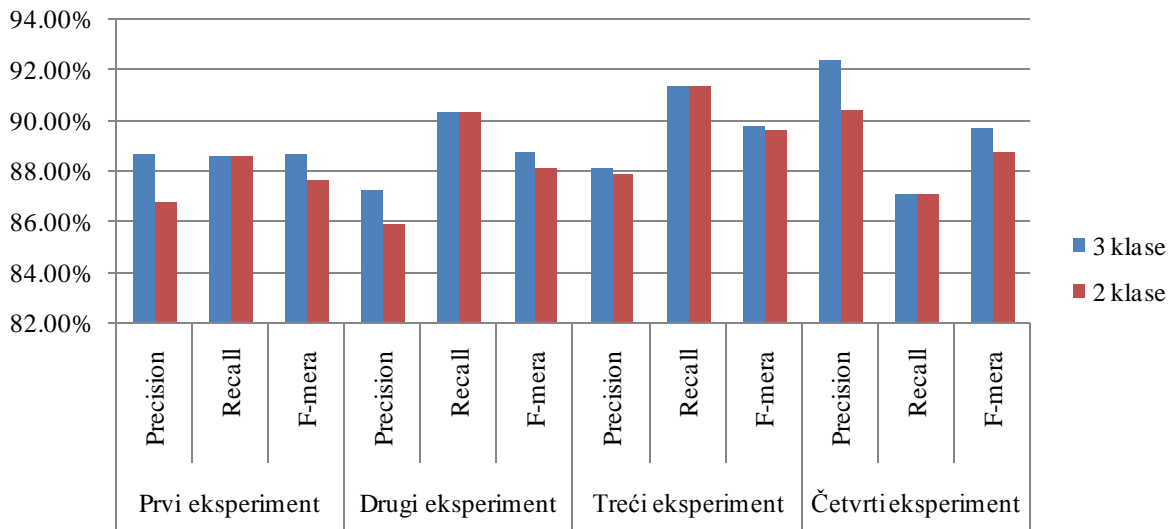
### Analiza uticaja neutralne klase na kvalitet klasifikacije

Može se zapaziti da je neutralna klasa loše klasifikovana u svim razvijenim modelima, što može biti posledica nezastupljenosti ove kategorije komentara u korpusu i činjenice da anotacija sentiment reči nije vršena za neutralne pojmove.

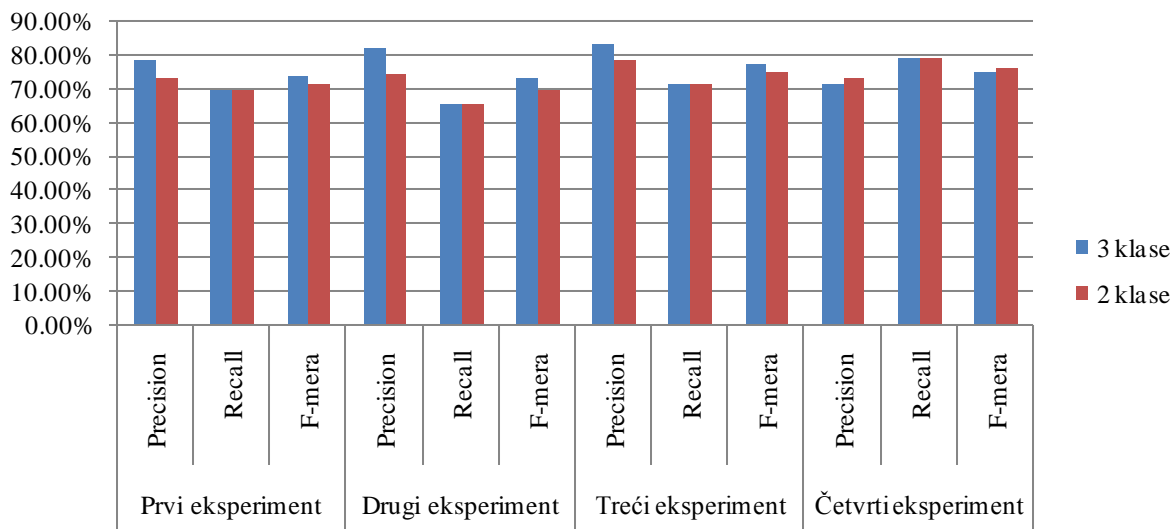
Da bi se ispitalo uticaj neutralne klase na kvalitet klasifikacije, reprodukovani su goreprikazani i opisani eksperimenti na nivou dokumenta i na nivou rečenice, s tim da su upotrebljena dva skupa anotiranih podataka za obučavanje: klasa pozitivnih recenzija/rečenica i klasa negativnih

recenzija/rečenica, dok je neutralna klasa izostavljena. Evaluacija razvijenih modela je vršena nad podacima iz skupa za testiranje. Kako su ovi modeli poslužili za ilustraciju uticaja neutralne klase na klasifikaciju i nisu upotrebljeni u daljem istraživanju, detaljan prikaz matrica konfuzije sa evaluacionim merama klasifikacije dat je u [Prilogu L](#).

Slike 4.33. i 4.34. ilustruju oscilacije u kretanju vrednosti evaluacionih mera *klasifikacije na nivou dokumenta*. Plavom bojom su označene vrednosti razmatranih mera u klasifikaciji kojom je obuhvaćena neutralna klasa, dok su crvenom bojom označene vrednosti u klasifikaciji kojom nije obuhvaćena neutralna klasa.



Slika 4.33. Klasifikacija na nivou dokumenta: Paralelni prikaz vrednosti evaluacionih mera za pozitivnu klasu

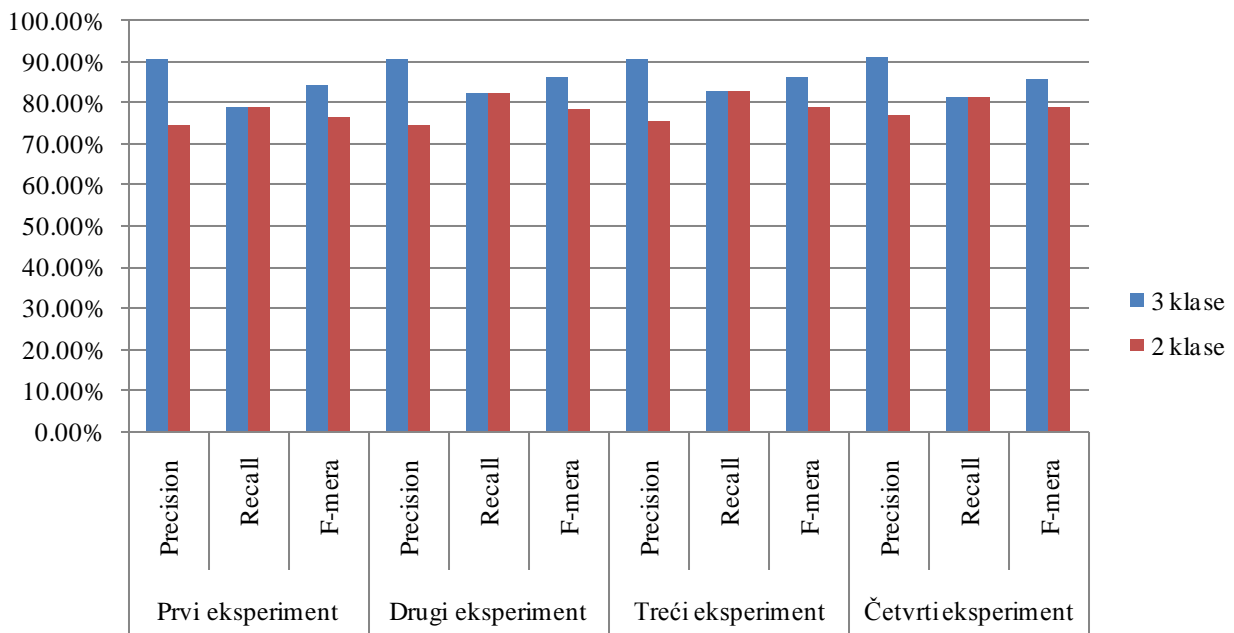


Slika 4.34. Klasifikacija na nivou dokumenta: Paralelni prikaz vrednosti evaluacionih mera za negativnu klasu

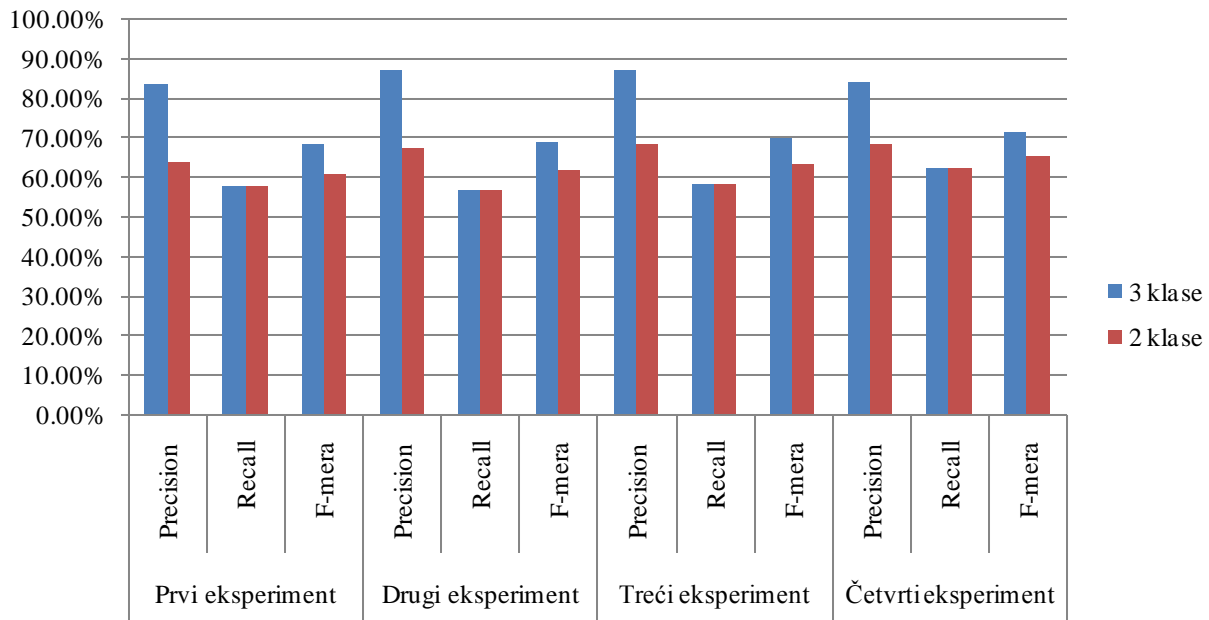
Na osnovu prikazanih podataka može se zaključiti sledeće:

- Bez obzira na posmatranu klasu, vrednosti parametra *Recall* ostaju nepromenjene, odnosno neutralna klasa ima uticaja samo na preciznost klasifikacije iskazanu merom *Precision*. Promene u vrednostima mere *Precision* dovode do promena u vrednosti F-mere.
- Veće oscilacije u vrednostima evaluacionih mera se zapažaju kod pozitivne klase. Vrednosti svih posmatranih mera značajno opadaju u uslovima kada klasifikacijom nije obuhvaćena neutralna klasa. Neki od potencijalnih uzroka ove pojave mogu biti: a) pojava novih sentiment reči pozitivne semantike u skupu za testiranje, koje nisu obuhvaćene sentiment rečnicima, b) nedostatak konteksta u kojem se pozitivne reči koriste, c) implicitno iskazivanje pozitivnog sentimenta (npr. u rečenici „Jedna je profesorka Mina“ nema nijedne eksplicitne sentiment reči i rečenica će biti okarakterisana kao neutralna). Budućim istraživanjima će se detaljnije ispitati razlozi ove pojave.
- U slučaju negativne klase zabeležen je neznatni pad vrednosti evaluacionih mera. Četvrti eksperiment predstavlja izuzetak – dolazi do blagog porasta vrednosti mere *Precision* i posledično F-mere.

Slike 4.35. i 4.36. ilustruju oscilacije u kretanju vrednosti evaluacionih mera *klasifikacije na nivou rečenice*. Plavom bojom su označene vrednosti razmatranih mera u klasifikaciji kojom je obuhvaćena neutralna klasa, dok su crvenom bojom označene vrednosti u klasifikaciji kojom nije obuhvaćena neutralna klasa.



Slika 4.35. Klasifikacija na nivou rečenice: Paralelni prikaz vrednosti evaluacionih mera za pozitivnu klasu



Slika 4.36. Klasifikacija na nivou rečenice: Paralelni prikaz vrednosti evaluacionih mera za negativnu klasu

Na osnovu prikazanih podataka može se zaključiti sledeće:

- Kao i u slučaju klasifikacije na nivou dokumenta, vrednosti mere *Recall* ostaju nepromenjene, odnosno nisu pod uticajem neutralne klase.
- Bez obzira na klasu koja se posmatra, vrednosti mere *Precision* značajno opadaju u uslovima kada klasifikacijom nije obuhvaćena neutralna klasa. U klasifikaciji na nivou rečenice ova pojava je izraženija kod negativne klase, što može biti posledica dužine analiziranog sadržaja i varijabilnosti u negativnim sentiment rečima. Buduća istraživanja baviće se detaljnijom analizom uzroka ove pojave.

Iz navedenog se zaključuje da neutralna klasa ima značajnu ulogu u klasifikaciji sentimenta studentskih recenzija. S obzirom da se ona identifikuje sa najmanje uspeha, bez obzira na nivo klasifikacije, cilj budućih istraživanja biće detaljnije modelovanje neutralne klase. Pretpostavka je da bi uvođenje rečnika neutralnih reči podiglo performanse rada modela po pitanju neutralne klase kada je reč o klasifikaciji i na nivou dokumenta i na nivou rečenice. Takođe, u prezentovanim modelima veliki problem predstavlja nezastupljenost neutralne klase. Usled nedostatka podataka, ovu klasu nije bilo moguće uspešno modelovati.

#### 4.1.6.3. Diskusija rezultata

Razvijeni korpus studentskih recenzija, opisan u [Poglavlju 4.1.3](#), koji je razdvojen na skup za obučavanje i skup za testiranje, upotrebljen je za razvoj modela sentiment analize primenom algoritama nadgledanog učenja i za razvoj modela sentiment analize zasnovane na leksikonima (skup za obučavanje), kao i za njihovu evaluaciju na osnovu skupa za testiranje. U oba slučaja klasifikacija je vršena na nivou recenzije i na nivou rečenice.

U [Poglavlju 4.1.5.1](#) i [Poglavlju 4.1.6.1](#) eksperimentisalo se algoritmima mašinskog učenja: Naïve Bayes, Support Vector Machines i k-Nearest Neighbor. U nastavku su sumirana ključna zapažanja i rezultati:

- Eksperimentisanje u fazi pripreme podataka pokazalo je da bigrami daju najbolje rezultate za naš korpus, bez obzira na nivo klasifikacije. Izuzetak je SVM klasifikator celih recenzija, čije su performanse najbolje kada se koriste unigrami, odnosno lista pojedinačnih reči koje se pojavljuju u skupu podataka. Superiornost bigrama u odnosu na druge kombinacije n-gramova je verovatno posledica veće količine podataka u skupu za obučavanje, kao i činjenice da studenti neretko koriste određene fraze ili specifičan sled reči u iskazivanju sentimenta. Kako je obično teško odrediti količinu podataka koja je dovoljna za obučavanje, buduće analize će pokazati da li su bigrami adekvatno rešenje i u uslovima kada se raspolože manjom količinom podataka.
- Na osnovu *TF-IDF* mere, opisane u [Poglavlju 3.2.2](#), postižu se dobri rezultati. Na nivou dokumenta, kNN algoritam predstavlja globalno najbolje rešenje. F-mera za pozitivnu klasu iznosi 88.94%, dok za negativnu klasu iznosi 74.28%. Na nivou rečenice SVM algoritam predstavlja globalno najbolje rešenje. F-mera za pozitivnu klasu iznosi 84.53%, a za negativnu 72.21%. Na osnovu dobijenih rezultata može se zaključiti da se primenom algoritama nadgledanog učenja postižu bolji rezultati na nivou dokumenta.
- Bez obzira na nivo klasifikacije, SVM i kNN manje uspešno klasifikaciju negativne recenzije, odnosno rečenice, dok Naïve Bayes algoritam sa manje uspeha klasifikuje pozitivne recenzije, odnosno rečenice. Ova činjenica povlači razlike u vrednostima mera *Precision* i *Recall* za klasu pozitivnih, odnosno negativnih recenzija/rečenica.
- Generalno posmatrano, svi razvijeni modeli iskazuju lošije performanse po pitanju identifikovanja negativne klase. Niže vrednosti mere *Recall* za negativnu klasu ukazuju na to da potencijalni problem mogu biti veće varijacije u vokabularu u negativnim recenzijama, te da unigramima i bigramima nisu potpuno obuhvaćene reči koje oslikavaju negativan sentiment. Takođe, unigrami i bigrami ne oslikavaju kontekst u kom su reči upotrebljene. Uključivanje informacije o kontekstu i sentiment statusu susednih rečenica može biti korisno za podizanje performansi klasifikatora. Ovi aspekti i njihov uticaj na performanse klasifikatora detaljnije će se analizirati u budućim istraživanjima.
- Kao što tabela 4.3., data u [Poglavlju 4.1.1](#), pokazuje, u korpusu je prisutno blizu 61% pozitivnih recenzija i blizu 38% negativnih. Iako ovakva distribucija nije atipična za

recenzije, nebalansiranost podataka takođe može biti uzrok promena u vrednostima evaluacionih mera klasifikatora. Buduća istraživanja baviće se analizom tačnosti klasifikacije u uslovima balansiranih podataka, te će se detaljnije ispitati njen uticaj na performanse klasifikatora.

- Može se pretpostaviti da bi se performanse klasifikatora poboljšale i uvođenjem atributa izvedenih na osnovu zapažanja prikazanih u [Poglavlju 4.1.3.3](#), što će takođe biti predmet budućih istraživanja.

Klasifikacija sentimenta na osnovu razvijenih rečnika opisana je u [Poglavlju 4.1.5.2](#). Sproveden je niz eksperimenata u kojima se analizirala upotreba različitih rečnika i njihov efekat na rezultate klasifikacije, a evaluacija razvijenih modela prikazana je u [Poglavlju 4.1.6.2](#). U nastavku su sumirana ključna zapažanja i rezultati:

- Uvođenjem domen-specifičnih rečnika pospešuje se preciznost klasifikacije i na nivou rečenice i na nivou dokumenta.
- Rezultati su ukazali na značaj manuelne evaluacije sentiment rečnika. Nakon korekcije rečnika pozitivnih i negativnih sentiment reči, koja je podrazumevala uklanjanje reči koje ne nose sentiment, kao i reči koje su pogrešnim sentimentom označene, postiže se preciznije identifikovanje negativnih recenzija – vrednost mere *Precision* raste za približno 3.50%. Po pitanju pozitivne klase nisu zabeleženi značajniji efekti.
- Uvođenje rečnika intenzifikatora, neutralizatora i rečnika reči koje menjaju sentiment orijentaciju naredne reči u rečenici donosi suptilne pomake u vrednostima evaluacionih mera klasifikatora. Može se pretpostaviti da se razlog minornog efekta, koji uvođenje dodatnih rečnika ima na tačnost klasifikacije, ogleda u broju upotrebljenih sentiment reči u jednoj recenziji/rečenici. Studenti u proseku koriste 2 pozitivne reči u pozitivnim recenzijama, odnosno 1 negativnu reč u negativnim recenzijama, kao što pokazuju podaci u tabeli 4.15., koja je data u [Poglavlju 4.1.3.3](#). One se identifikuju na osnovu rečnika pozitivnih i negativnih reči. Otuda se suptilno povećanje u vrednostima evaluacionih mera može prevashodno pripisati rečniku reči koje menjaju sentiment orijentaciju naredne reči u rečenici. Intenzifikatori i neutralizatori jačaju i slabe intenzitet iskazanog sentimenta i kao takvi mogu imati ograničen uticaj na preciznost klasifikacije – intenzifikatori i neutralizatori mogu doneti prevagu samo u slučaju vrednosti sentimenta koja teži nuli.
- Pristup klasifikaciji sentimenta zasnovan na leksikonima daje dobre rezultate. U klasifikaciji na nivou dokumenta postignuta je F-mera za pozitivnu klasu 89.75%, a za negativnu 77.28%. U klasifikaciji na nivou rečenice F-mera za pozitivnu klasu iznosi 86.05%, a za negativnu 71.80%. U oba slučaja date su vrednosti F-mere eksperimenata koji daju najbolje rezultate na različitim nivoima klasifikacije, a koji ujedno uključuju sve gorepomenute rečnike.
- Rezultati klasifikacije pokazuju da se vrednost mere *Recall* za negativnu klasu značajno smanjuje kada se modeli primene nad test podacima. Ovo je posebno izraženo u



klasifikaciji na nivou rečenice. Kako je već istaknuto, studenti u iskazivanju negativnog stava koriste bogatiji rečnik, u odnosu na rečnik kojeg koriste u iskazivanju pozitivnog stava. Prema našem mišljenju, značajno smanjenje vrednosti mere *Recall* za negativnu klasu može biti posledica pojave sentiment reči ili specifičnih fraza u skupu za testiranje, koje se nisu pojavile u skupu za obučavanje, te iz tog razloga nisu obuhvaćene sentiment rečnicima. Predmet budućih istraživanja biće detaljna analiza uzroka pojave ove greške. Prednost modela zasnovanog na rečnicima se ogleda u mogućnosti jednostavnog i brzog proširenja sentiment rečnika kako bi se obuhvatili nedostajući izrazi.

- Svi razvijeni modeli neuspešno vrše klasifikaciju neutralne klase, što može biti posledica nezastupljenosti ove kategorije komentara u korpusu i činjenice da anotacija sentiment reči nije vršena za neutralnu klasu. Buduća istraživanja baviće se problemom nebalansiranosti podataka i razmatranjem mogućnosti efikasnijeg razdvajanja subjektivnih recenzija od onih u kojima su iskazane činjenice (bez sentimenta), a koje predstavljaju neutralnu klasu u sprovedenom istraživanju.

#### Generalni zaključci:

- Klasifikacija sentimenta u studentskim recenzijama realizovana je uspešnije primenom klasifikacije zasnovane na leksikonima, u odnosu na klasifikaciju primenom algoritama nadgledanog učenja. U najboljim modelima na nivou dokumenta F-mera za pozitivnu klasu je viša za 0.81%, dok je za negativnu klasu viša za 3%. Na nivou rečenice F-mera za pozitivnu klasu je viša za 1.52%. Izuzetak predstavlja klasifikacija negativne klase na nivou rečenice, koja se nešto uspešnije identifikuje primenom algoritama mašinskog učenja (u 0.41% slučajeva).
- S obzirom da studenti u jednoj recenziji, odnosno jednoj rečenici, evaluiraju više različitih aspekata i da se reči pozitivne semantike, odnosno negativne, koriste u podjednako meri u pohvalama i pokudama različitih aspekata poslovanja visokoškolskih institucija razlikovanje aspekata samo na osnovu sentiment reči ne bi bila moguća. Buduća istraživanja će se baviti iznalaženjem adekvatnih atributa za klasifikaciju sentimenta prema aspektima.

U članku (Mullich, 2012) je istaknuto da tačnost klasifikacije ne predstavlja najbitniji parametar evaluacije modela sentiment analize. Poslovni problem koji se rešava sentiment analizom određuje parametar relevantan za datu aplikaciju sentiment analize. Primena sentiment analize u borbi protiv terorizma zahtevala bi sistem koji ima vrednost *Recall* mere 100%, dok bi sistem mogao da toleriše nisku preciznost, navodeći nadalje da je za mnoge aplikacije sentiment analize za potrebe marketinga dovoljna tačnost klasifikacije od 70%. Za potrebe identifikovanja sentimenta u komentarima o visokoškolskim institucijama i profesorima relevantniji parametar je preciznost klasifikacije (npr. identifikovanje rečenice *Profesor ne radi svoj posao korektno* kao pozitivne, ne može se tolerisati). U većini eksperimenata razvijeni modeli pokazuju veću

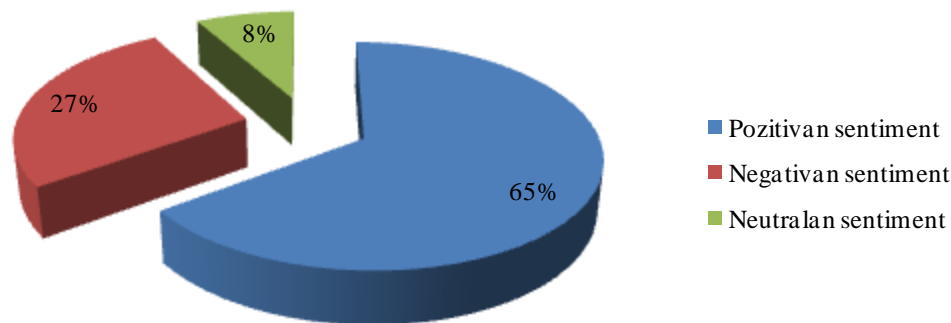
vrednost mere *Precision*, u odnosu na vrednost mere *Recall*. Imajući na umu navedeno, može se zaključiti da razvijeni klasifikatori imaju zadovoljavajuću preciznost rada.

## **4.2. Mogućnost unapređenja poslovanja visokoškolskih institucija primenom zaključaka analize**

Kreiranje evaluacionog korpusa, kao osnovnog preduslova za sprovođenje sentiment analize, i konsekvantna analiza pripremljenog sadržaja predstavljali su neophodne korake za testiranje polazne hipoteze. Na ovaj način stvoreni su uslovi za ispitivanje mogućnosti upotrebe deriviranih saznanja u poboljšanju odnosa visokoškolske institucije i njenih studenata kroz građenje lojalnosti postojećih studenata, negovanje reputacije i prepoznatljivosti brenda, kao i prilagođavanje usluga preferencijama studenata i tržišnim trendovima.

Iskustvo studenata se može klasifikovati u jednu od tri osnovne kategorije: pozitivno, negativno i neutralno iskustvo. Upotrebom sentiment analize visokoškolske institucije mogu detektovati ton i temperament svake reči u studentskim komentarima i kategorizovati te sentimente kao pozitivne, negativne ili neutralne. Na osnovu dobijenih saznanja visokoškolske institucije stiču vredan uvid o onome što studenti smatraju ispravnim poslovanjem (pozitivan sentiment) i o aspektima poslovanja koje bi, prema njihovom mišljenju, trebalo unaprediti (negativan sentiment). Stavljanjem ovih saznanja u funkciju poslovanja unapređuje se sveukupno iskustvo i zadovoljstvo studenata.

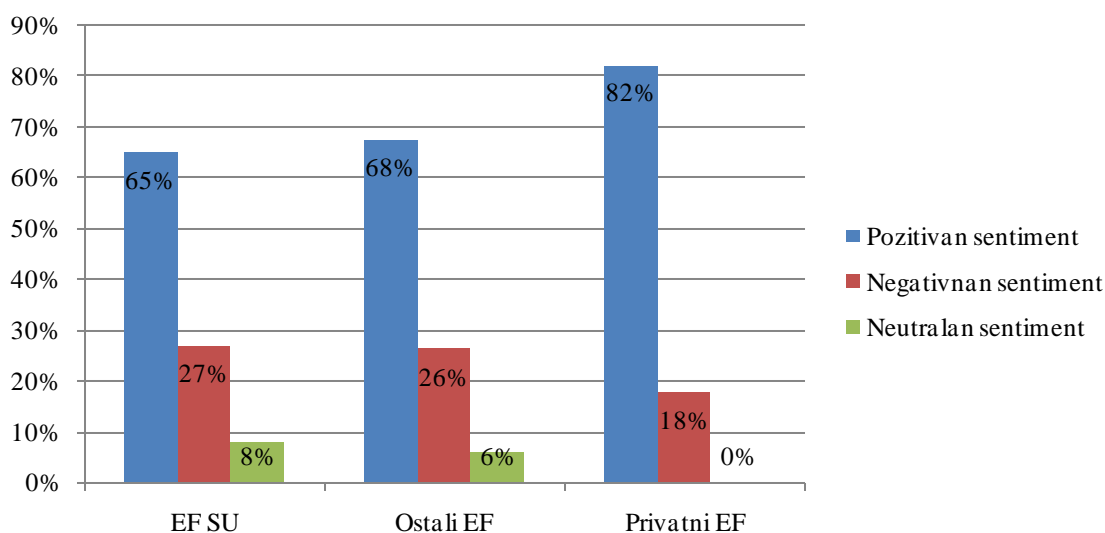
Na primeru Ekonomskog fakulteta u Subotici ilustrovane su mogućnosti upotrebe razvijenih modela prikazanih u prethodnom poglavlju. Studenti Ekonomskog fakulteta u Subotici, koji su načinili 4.743 ocena profesora, najaktivniji su članovi veb sajta *Oceni profesora*. Međutim, pri ocenjivanju studenti nisu u obavezi da ostave komentar. Otuda je za ovaj fakultet prikupljeno 468 studentskih recenzija. Slika 4.37. prikazuje njihovu strukturu prema polaritetu iskazanog sentimenta. Studenti pretežno iskazuju naklonost prema nastavnom osoblju Ekonomskog fakulteta u Subotici: sentiment je u 65% komentara pozitivan, a u 27% negativan.



**Slika 4.37. Učešće pojedinih kategorija sentimenta u recenzijama nastavnog osoblja Ekonomskog fakulteta u Subotici**

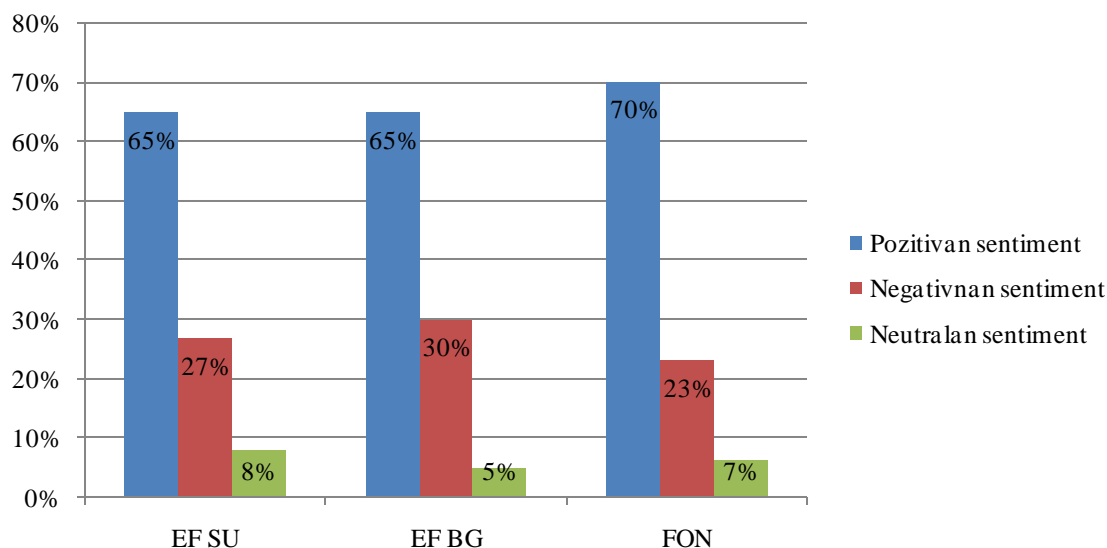
Na osnovu prikazanih podataka moglo bi se zaključiti da je zadovoljstvo studenata pretežno zadovoljavajuće i da fakultet pruža adekvatan nivo obrazovanja, s obzirom na to da većina studenata iskazuje pozitivan stav. Međutim, samo poređenjem sopstvene onlajn reputacije sa onlajn reputacijom konkurencije visokoškolske institucije mogu sagledati svoju poziciju u odnosu na konkurenciju i steći realnu sliku o sentimentu studenata i onlajn reputaciji institucije. Saznanje o raspoloženju studenata ka konkurentima pomaže visokoškolskim institucijama da procene sopstveni rad i pronađu načine za unapređenje.

Slika 4.38. prikazuje poređenje onlajn reputacije Ekonomskog fakulteta u Subotici i ostalih državnih i privatnih Ekonomskih fakulteta. Podaci ilustruju da je nezadovoljstvo studenata za 1% iznad nezadovoljstva studenata na ostalim državnim ekonomskim fakultetima, dok je znatno iznad nezadovoljstva studenata koji studiraju na privatnim fakultetima – za 9%. Ovakvo globalno poređenje govori da je onlajn reputacija Ekonomskog fakulteta u Subotici približna reputaciji ostalih državnih ekonomskih fakulteta.



Slika 4.38. Poređenje onlajn reputacije Ekonomskog fakulteta u Subotici i ostalih Ekonomskih fakulteta u Srbiji

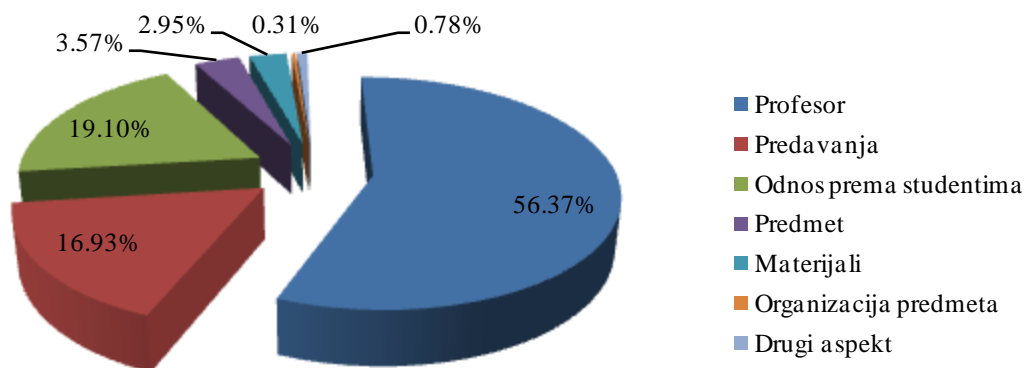
U skladu sa raspoloživim podacima prikupljenih sa sajta *Oceni profesora* detaljnija komparacija sentimenta je izvršena između Ekonomskog fakulteta u Subotici i dva državna fakulteta iz domena ekonomije i poslovanja. To su Ekonomski fakultet u Beogradu i Fakultet organizacionih nauka. Recenzije o profesorima koji predaju na drugim konkurentskim fakultetima nisu postojale na sajtu *Oceni profesora*, ili ih je bilo suviše malo usled neaktivnosti njihovih studenata. Rezultati poređenja onlajn reputacije prikazani su na slici 4.39.



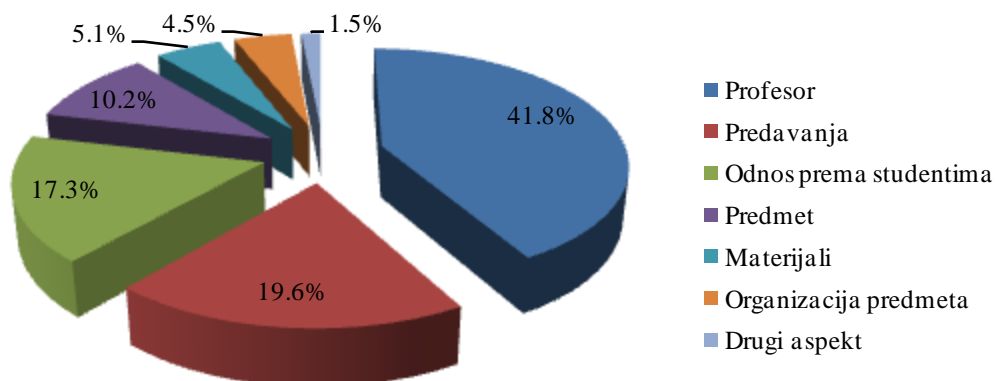
Slika 4.39. Komparacija onlajn reputacije po institucijama

Može se zaključiti da je Ekonomski fakultet u Subotici bolje pozicioniran kod svojih studenata u odnosu na Ekonomski fakultet u Beogradu, gde je 30% komentara negativno. U poređenju sa studentima Ekonomskog fakulteta u Subotici, studenti Fakulteta organizacionih nauka su zadovoljniji, te je procenat negativnih kritika niži za 4%. Iako onlajn reputacija Ekonomskog fakulteta u Subotici ne odstupa značajno od ostalih institucija, činjenica da 27% komentara nosi negativnu semantiku ukazuje na to da postoje potrebe za poboljšanjima u radu nastavnog osoblja i da ovaj procenat nije u potpunosti prihvatljiv.

Anotacija korpusa je vršena za aspekte, odnosno predmet recenziranja, i za sentiment koji studenti iskazuju prema recenziranom aspektu. Ove informacije su upotrebljene u daljoj analizi. Slike 4.40. i 4.41. prikazuju distribuciju aspekata u pozitivnim i negativnim recenzijama nastavnog osoblja Ekonomskog fakulteta u Subotici, respektivno. Osnovna svrha sajta *Oceni profesora* jeste da se izvrši evaluacija nastavnog osoblja, otuda je prirodno da je najveći procenat pozitivnih i negativnih recenzija upravo za aspekt *Profesor*, koji se odnosi na uopštene karakteristike profesora kao osobe i kao predavača. Studenti pozitivno ocenjuju *Odnos prema studentima*, zatim *Predavanja*. U negativnim recenzijama, koje su prikazane na slici 4.41, osim komentara koji se tiču aspekta *Profesor*, dominiraju komentari o aspektu *Predavanja*.



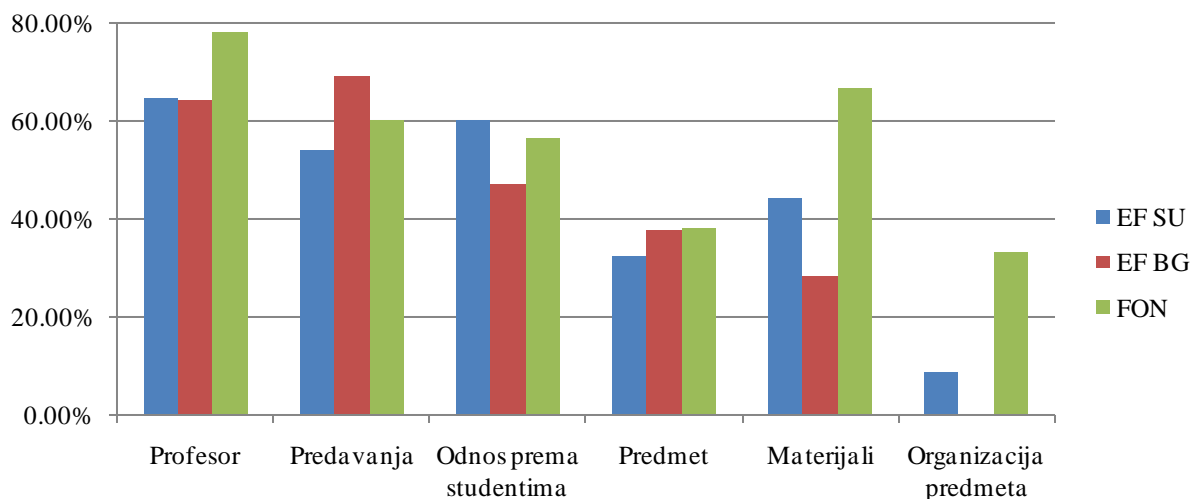
Slika 4.40. Distribucija pozitivnih komentara po aspektima – primer Ekonomskog fakulteta u Subotici



Slika 4.41. Distribucija negativnih recenzija po aspektima – primer Ekonomskog fakulteta u Subotici

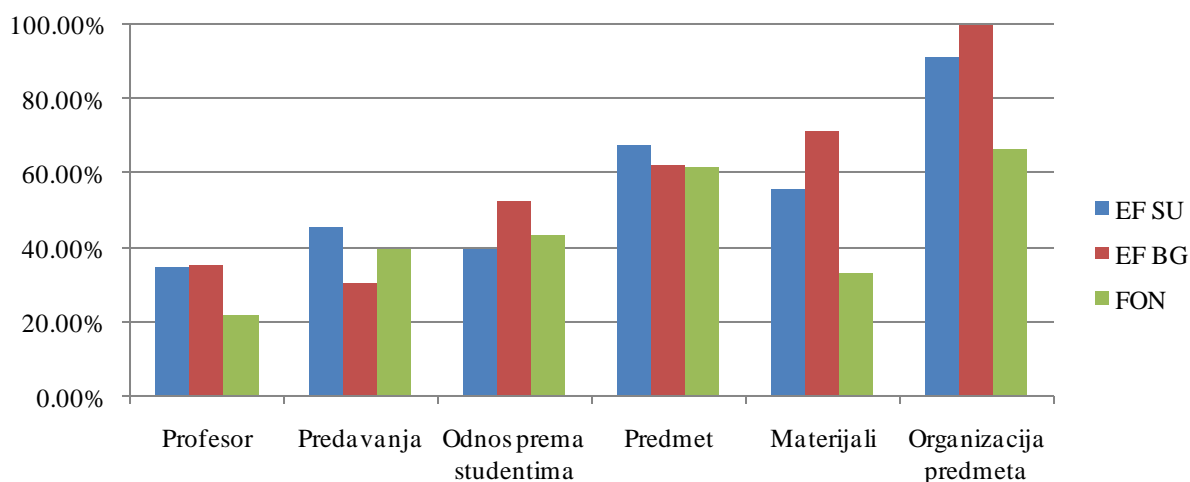
Kao i kod analize generalnog sentimenta studenata, potrebno je sagledati poziciju Ekonomskog fakulteta u odnosu na konkurenciju kako bi se izvršila sveobuhvatna procena stepena (ne)zadovoljstva studenata i prihvatljivost iskazanih procenata.

Slika 4.42. daje paralelni prikaz kretanja pozitivnih recenzija po aspektima i institucijama. Fakultet organizacionih nauka prednjači po pozitivnim ocenama aspekata *Profesor*, *Materijali* i *Organizacija predmeta*. Ekonomski fakultet u Subotici je u boljoj poziciji u odnosu na konkurenciju jedino po pitanju ocene *Odnosa profesora prema studentima*, dok je lošije pozicioniran od dve posmatrane institucije po pitanju zadovoljstva studenata aspektom *Predmet*. Ekonomski fakultet u Beogradu prednjači po zadovoljstvu studenata u pogledu *Predavanja*.



Slika 4.42. Kretanje pozitivnih recenzija prema aspektima

Iz ugla poslovanja, relevantnije su negativne recenzije, koje ukazuju na nezadovoljstvo studenata i uska grla u poslovanju na kojima je potrebno raditi radi unapređenja odnosa sa studentima. Na osnovu ilustracije prikazane na slici 4.43. može se zaključiti da je Ekonomski fakultet u Subotici lošije pozicioniran u odnosu na ostale posmatrane institucije po pitanju kvaliteta predavanja i sadržine predmeta. Studenti Ekonomskog fakulteta u Beogradu intenzivnije negativno komentarišu odnos profesora prema studentima, nastavne materijale i udžbenike, kao i organizaciju predmeta.

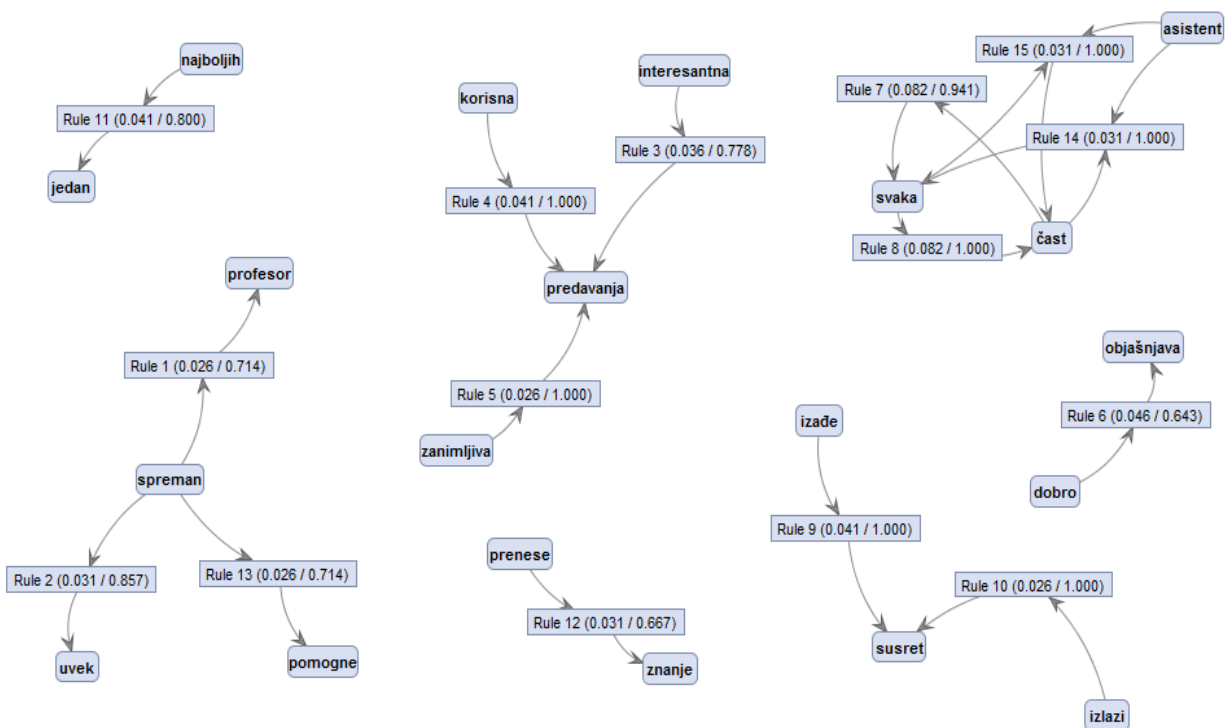


Slika 4.43. Kretanje negativnih recenzija prema aspektima

Asocijativna pravila predstavljaju tehniku *data mining*-a koja se uspešno koristi u determinisanju ponašanja potrošača identifikovanjem pravila koja ukazuju na učestale paterne, korelacije i

zavisnosti u podacima. Evaluacija rezultujućih pravila vrši se na osnovu dva parametra: signifikantnost pravila (engl. *support*) i pouzdanost pravila ili stepen poverenja (engl. *confidence*). U daljoj analizi asocijativna pravila su primenjena nad skupovima pozitivnih i negativnih recenzija s ciljem identifikovanja reči koje se učestalo koriste zajedno i oblikovanja zadovoljstva i nezadovoljstva studenata. Pravila su generisana u RapidMiner alatu. Alat svakom pravilu automatski dodeljuje nazive (*Rule 1, Rule 2, itd.*) nakon kojih slede vrednosti dva parametra, respektivno.

Analizom asocijativnih pravila u pozitivnim i pretežno pozitivnim recenzijama o nastavnom osoblju Ekonomskog fakulteta u Subotici ustanovljeno je da se u pozitivnim recenzijama izdvaja sedam grupacija asocijativnih pravila prikazanih na slici 4.44. Izdvojena su pravila sa stepenom poverenja većim od 0.60, što govori da je pravilo tačno u 60% ili više slučajeva u kojima se premisa pravila pojavila. Na primer: Ukoliko se u recenziji pojavi reč *interesantna*, u 77.8% slučajeva ona se odnosi na predavanja, dok se reči *korisna* i *zanimljiva* uvek odnose na predavanja, jer je stepen poverenja za dato pravilo 1. Na osnovu generisanih asocijativnih pravila može se zaključiti da studenti Ekonomskog fakulteta u Subotici cene profesore koji su spremni da im pomognu i izađu u susret, koji znaju da objasne gradivo i prenesu znanje, kao i predavanja koja su interesantna, korisna i zanimljiva.



Slika 4.44. Asocijativna pravila u pozitivnim recenzijama nastavnog osoblja Ekonomskog fakulteta u Subotici

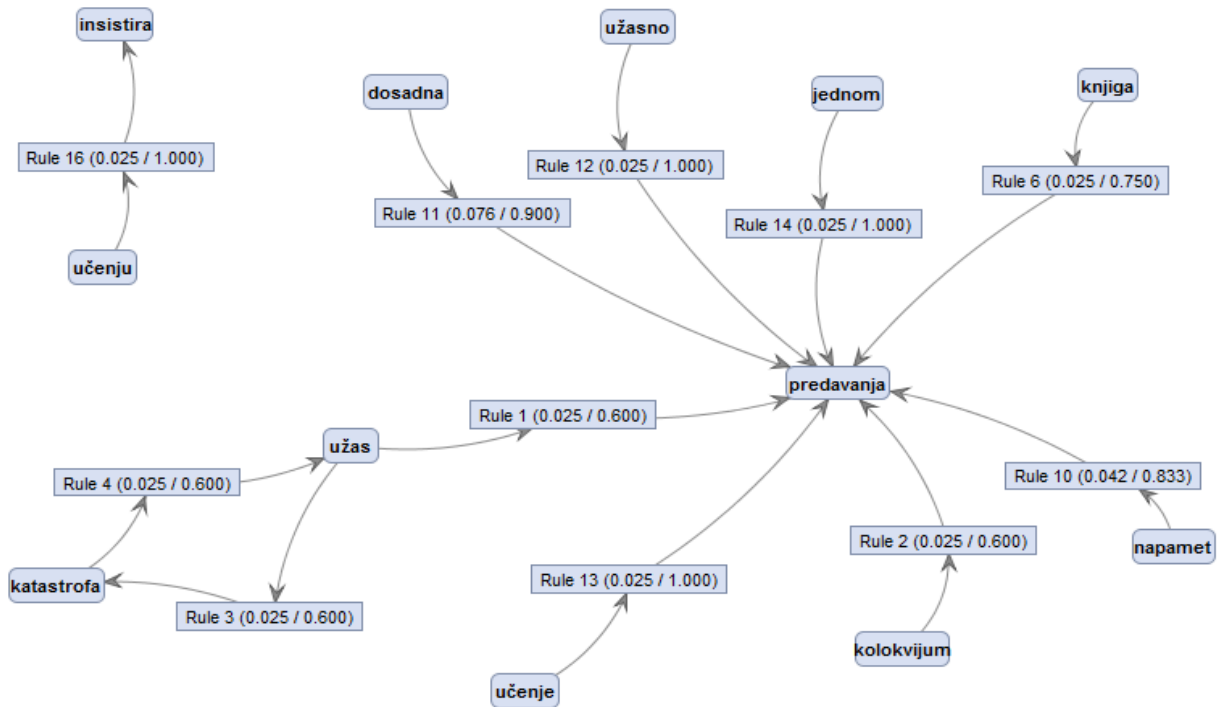
Detaljnije sagledavanje komentara zahteva analizu i otkrivanje važnih fraza koje mogu i koje bi trebalo da utiču na oblikovanje usluge, podrške studentima, marketing poruke fakulteta i dometa koji fakultet ostvaruje putem društvenih medija. Bigrami i trigrami daju dve, odnosno tri uzastopne reči (frazе) koje se frekventno pojavljuju u skupu. Njihovom analizom u pozitivnim i pretežno pozitivnim recenzijama o nastavnom osoblju Ekonomskog fakulteta u Subotici identifikovan je niz fraza koje jasno ukazuju na pozitivan sentiment, ali nisu dovoljno informativne. Primeri ovakvih fraza su: *najbolji profesor/asistent, izuzetan predavač, odličan profesor/asistent, divan čovek / divna žena*. Ovim i sličnim frazama studenti iskazuju zadovoljstvo, ali ne opisuju razloge koji profesora čine najboljim i izuzetnim u svom poslu. Navedene fraze su dominantne u skupu pozitivnih recenzija, međutim, prisutne su i druge, informativnije, koje su sažete u nastavku.

Analizom asocijativnih pravila i značajnih fraza identifikovanih u skupu pozitivnih recenzija, zadovoljstvo studenata Ekonomskog fakulteta u Subotici može se sumirati sledećim iskazima:

- Profesor dobro objašnjava nastavno gradivo i ume da prenese znanje.
- Predavanja su interesantna.
- Profesor je korektan prema studentima, izlazi im u susret kada je potrebno, dostupan je studentima, ceni znanje i ne zahteva učenje od reči do reči.
- Profesor je elokventan, stručan, raspolaže velikim znanjem i profesionalno obavlja radne zadatke.

Analizom asocijativnih pravila u negativnim i pretežno negativnim recenzijama o nastavnom osoblju Ekonomskog fakulteta u Subotici ustanovljeno je da se u negativnim recenzijama pretežno izdvajaju pravila povezana sa kvalitetom predavanja. Asocijativna pravila, identifikovana u isključivo negativnim recenzijama u kojima je stepen poverenja veći od 0.60, prikazana su na slici 4.45. Sa ilustracije su uklonjena pravila koja nisu informativna. Može se zaključiti sledeće: ukoliko se u recenziji pojavi reč *dosadna*, u 90% slučajeva ona se odnosi na predavanja, dok se reč *užasno* uvek odnosi na predavanja, jer je stepen poverenja za dato pravilo 1. Na osnovu generisanih asocijativnih pravila može se izvesti zaključak da studenti Ekonomskog fakulteta u Subotici nisu zadovoljni predavanjima koja su neinteresantna i profesorima koji insistiraju na učenju „od reči do reči”.





Slika 4.45. Asocijativna pravila u negativnim recenzijama nastavnog osoblja Ekonomskog fakulteta u Subotici

Analizom bigrama i trigrama ustanovljeno je da u negativnim i pretežno negativnim recenzijama dominiraju fraze poput: *užasno dosadna (predavanja)*, *veoma dosadna, izuzetno dosadna (predavanja)*, *najdosadnija predavanja*, *dosadna predavanja*, *nezanimljiva predavanja*, *totalno nezainteresovan*, *učenje napamet*, *(pravi) primer rasula* i druge fraze. Dakle, u osnovi negativnih kritika leži problem kvaliteta i interesantnosti predavanja. Analizom asocijativnih pravila i značajnih fraza identifikovanih u skupu negativnih recenzija, nezadovoljstvo studenata Ekonomskog fakulteta u Subotici može se sumirati sledećim iskazima:

- Predavanja su izuzetno neinteresantna, nezanimljiva i zastarela.
- Profesori ne znaju kako da zainteresuju studente i održe im pažnju.
- Profesori ne prenose znanje kvalitetno i nezainteresovani su.
- Profesori insistiraju na učenju „od reči do reči“.

Analiza sentimenta studenata predstavlja polaznu tačku u ispitivanju nezadovoljstva studenata, nedostataka u poslovanju visokoškolske ustanove i postavljanju smernica za unapređenje rada institucije. Međutim, identifikovanje dominantnih fraza u negativnim komentarima samo ukazuje na postojanje problema, te je nakon otkrivanja dominantnih fraza potrebno dublje istražiti problematiku kako bi se preciznije oblikovao problem sa kojim se institucija susreće. Prema dostupnim i analiziranim sadržajima ustanovljeno je da je veći deo nezadovoljstva studenata usmeren ka predavanjima, njihovom kvalitetu i stilu predavanja profesora, kao i ka problemu

učenja „od reči do reči“, čime se ne podstiče aktivno razmišljanje i angažovanje studenata nego puka interpretacija sadržaja. Na osnovu anotiranog i analiziranog sadržaja omogućeno je izdavanje podataka o konkretnim profesorima koji su predmet negativnih recenzija. Iako ovakva analiza izlazi iz naučnih okvira doktorata, ona bi omogućila fokusiranje na problem i identifikovanje uskog grla. Mogućnosti rešenja navedenih problema se svakako odnose na kontinuelna stručna usavršavanja profesora, kojima bi se podigao kvalitet znanja i održala njihova savremenost, kao i kroz dodatna predavanja i obuke o savremenim i kreativnim tehnikama predavanja s obzirom na to da nastavno osoblje fakulteta tokom svog školovanja nije izučavalo metodiku nastave ni pedagogiju.

Analogno prethodnom pristupu analizirane su recenzije studenata o profesorima koji predaju na Ekonomskom fakultetu u Beogradu i na Fakultetu organizacionih nauka. Ustanovljeno je da studenti Ekonomskog fakulteta u Beogradu cene:

- Profesore koji tokom predavanja kvalitetno prenose znanje, što studentima olakšava pripremu ispita.
- Profesore koji studente podstiču na razmišljanje.
- Stručnost profesora.
- Dobrotu i korektnost profesora, kao i one profesore koji izlaze u susret studentima.

Studenti Fakulteta organizacionih nauka cene:

- Profesore koji imaju sposobnost da održe pažnju studentima tokom predavanja i čija predavanja su interesantna.
- Profesore koji su uvek spremni da izađu u susret studentima i koji su korektni u komunikaciji i uopšteno u odnosu i stavu prema studentima.
- Predavanja na kojima je gradivo dobro objašnjeno i znanje uspešno preneto.

Studenti Ekonomskog fakulteta u Beogradu kritikuju:

- Loše napisane nastavne materijale i udžbenike.
- Profesore koji ne uspevaju da prenesu znanje i ne održavaju pažnju studenata tokom nastave.
- Loše pedagoške sposobnosti profesora i neiskazivanje ljudskih kvaliteta u odnosima sa studentima.

Studenti Fakulteta organizacionih nauka kritikuju:

- Nekorektnost profesora u odnosima prema studentima.
- Uobražen stav profesora u komunikaciji sa studentima.

Na osnovu pohvala i kritika upućenih profesorima konkurentskih fakulteta može se zaključiti da, za razliku od studenata Ekonomskog fakulteta u Beogradu, studenti Ekonomskog fakulteta u

Subotici vrlo retko komentarišu loš kvalitet nastavnih materijala i udžbenika. Jedan od osnovnih izvora nezadovoljstva studenata Ekonomskog fakulteta u Beogradu jesu loše pedagoške sposobnosti profesora, a studenata Fakulteta organizacionih nauka – nekorektnost u komunikaciji. Kako navedeni aspekti kvaliteta rada nastavnog osoblja i odnosa prema studentima nisu dominantni u negativnim komentarima Ekonomskog fakulteta u Subotici, nego predstavljaju izolovane slučajeve ili se ne pojavljuju u skupu negativnih recenzija, oni se mogu tretirati kao prednost. Aspekti na kojima treba raditi, koji su identifikovani kao slabost fakulteta, a prednost konkurencije, odnose se na predavanja. Predavanja moraju biti interaktivna, a profesori se moraju obučiti savremenijim tehnikama predavanja kako bi se predavanja učinila interesantnim i kako bi se podstaklo aktivno angažovanje studenata i podsticala njihova kreativnost i razmišljanje.

Na portalu *Oceni profesora* studenti anonimno ostavljaju komentare, te ne postoji mogućnost identifikovanja uticajnih autora u zajednici. U analizi sadržaja sa društvenih mreža, poput *Facebook*-a i *Twitter*-a, i sa foruma, gde svaki autor ostavlja komentar pod jedinstvenim imenom, korisno bi bilo identifikovati uticajne pojedince koji podstiču negativnu komunikaciju da bi se neutralizovali komentari koji narušavaju reputaciju institucije. Identifikovanjem komentara koje ostavljaju uticajni autori na društvenom mediju koji se prati i praćenjem dometa koji taj komentar ima može se otkriti i broj ljudi na koje će sadržaj imati uticaja. Ukoliko poruke pojedinca imaju velik domet na društvenim medijima, efikasan način razrešenja problema bilo bi javno obraćanje na blogu, forumu ili drugom društvenom mediju, uz obećanje da će se institucija posvetiti datom problemu. Dodatno, direktnim obraćanjem uticajnom korisniku društvenih medija povodom identifikovanog problema i diskusijom o istom steklo bi se saznanje o mogućnostima razrešenja problema. Uticajni korisnici se ne smeju zanemariti i potrebno je tokom vremena pratiti njihove poruke i progres u odnosu. Objedinjavanje analitika o uticajnim korisnicima i sentiment analize podrazumeva kreiranje profila za svakog značajnog korisnika, putem kojeg se prate njihovi sentiment indeksi. Ukoliko određena osoba sve aspekte<sup>47</sup> negativno procenjuje u podjednako meri, njihov negativan sentiment se donekle može tretirati kao odstupanje (engl. *outlier*) i odstraniti iz skupa negativnih komentara koji se razmatraju. Vođenje profila korisnika koji obuhvata njihovo učešće na različitim platformama društvenih medija omogućava institucijama da prate naklonost i sentiment ka brendu kroz vreme, da prate namere pojedinaca, kao i njihov uticaj na odluke drugih. Navedene analize izlaze iz okvira doktorske disertacije i predstavljaju jedan od mogućih pravaca budućih istraživanja.

Intenzitet iskazanih osećanja predstavlja takođe važan aspekt analize sentimenta. Iz ugla visokoškolskih institucija, komentar: „Pokajao/la sam se što sam upisao/la ovaj fakultet. Nikako ga ne preporučujem!“, čime student ukazuje na nezadovoljstvo celokupnim studentskim iskustvom, nema podjednaku težinu kao komentar: „Predavanja na ovom predmetu nisu interesantna, odaberite drugi predmet!“, čime student ukazuje na nezadovoljstvo konkretnim

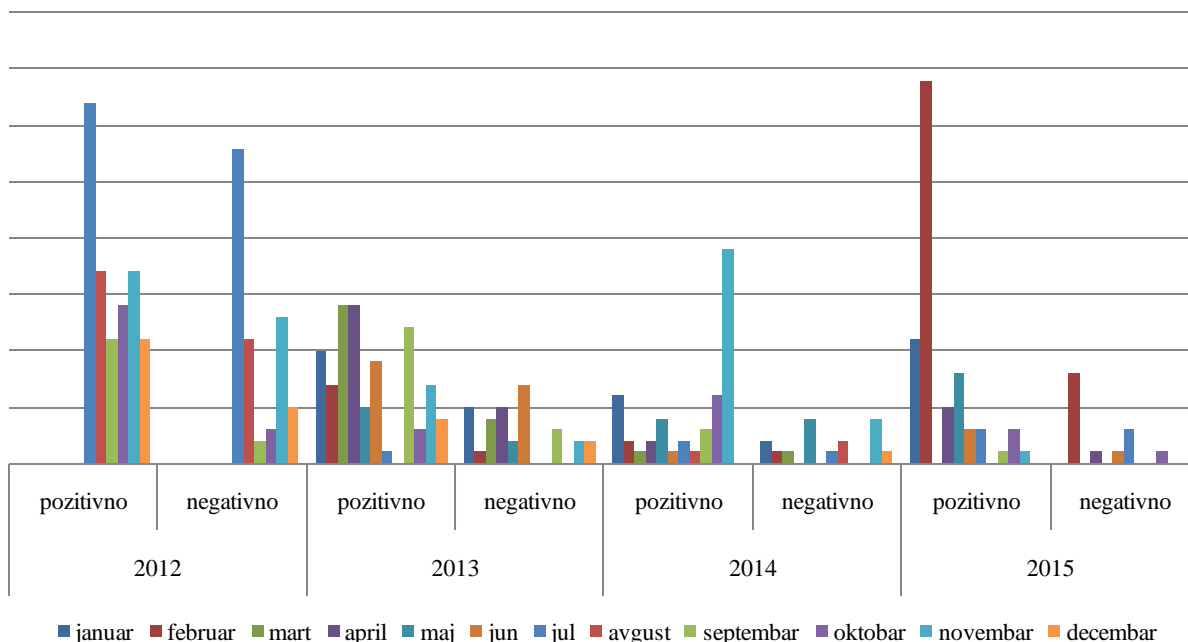
---

<sup>47</sup> U navedenom kontekstu „aspekt“ se odnosi na predmet recenziranja (proizvod ili njegove pojedinačne karakteristike, usluga ili njene karakteristike, brend, kompanija) i treba se razlikovati od „aspekta“ koji je predmet anotacije u istraživanju.

predmetom. Paralelnom analizom pozitivnih i negativnih komentara tri institucije može se zaključiti da studenti Ekonomskog fakulteta u Subotici, za razliku od studenata Ekonomskog fakulteta u Beogradu i Fakulteta organizacionih nauka, intenzivnije iskazuju emocije u sadržajima i pozitivne i negativne semantike. U pozitivnim komentarima učestalije koriste intenzifikatore, na primer: *jako dobro (lepo) objašnjava, jako dobar profesor (asistent, stručnjak), izuzetno dobro napisana (knjiga), izuzetno korektan*, i drugo. Na osnovu stila pisanja pozitivnih recenzija može se zaključiti da je intenzivniji pozitivan sentiment studenata Ekonomskog fakulteta u Subotici u odnosu na studente konkurentskih fakulteta. Međutim, isto se može zapaziti i u negativnim komentarima, gde studenti oštro kritikuju predavanja upotrebljavajući fraze poput: *izuzetno dosadna predavanja, najdosadnija predavanja, užasno dosadna predavanja i veoma dosadna predavanja*. Upotrebom navedenih intenzifikatora iskazani negativni sentiment je značajniji u odnosu na situaciju kada bi studenti navodili samo da su predavanja neinteresantna. U negativnim komentarima studenata Ekonomskog fakulteta u Beogradu zapaženo je da se učestalo koristi fraza *jako loš* uz reči čovek, pedagog ili profesor. Anotacijom sadržaja obuhvaćena je i informacija o intenzitetu sentiment reči, što je otvorilo mogućnost inkorporiranja ovih informacija u buduće sisteme za identifikovanje sentiment polariteta, čime se očekuje da će se pouzdanost sistema dodatno pospešiti i da će se time uspešnije razdvajati intenzivniji negativni komentari od negativnih komentara prosečne jačine.

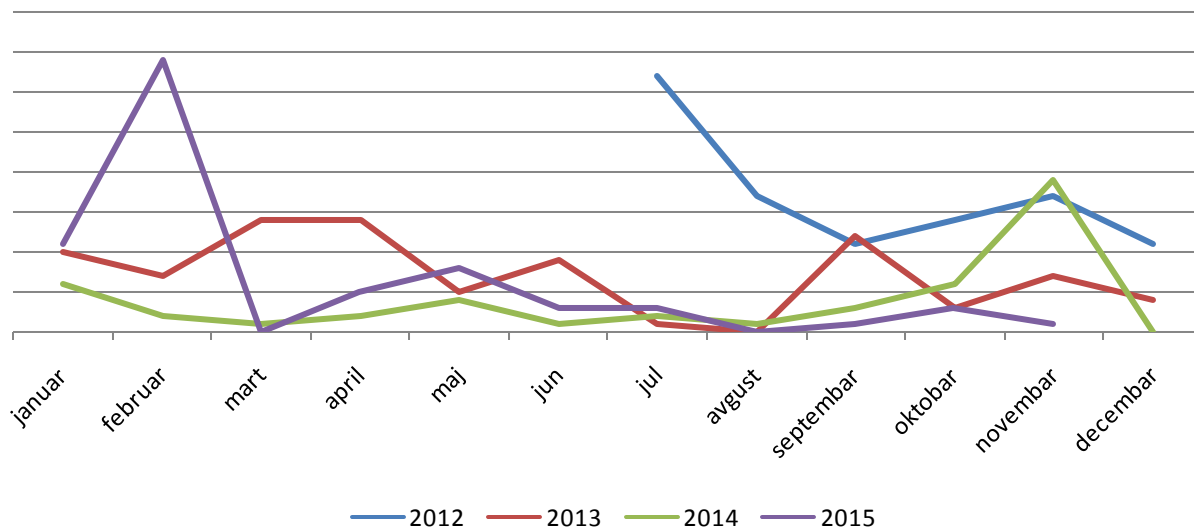
Sentiment analiza može pomoći u predviđanju trendova i preferencija i razvijanju strategije koja će omogućiti institucijama sticanje dobiti na osnovu identifikovanih trendova, kao i konkurentske prednosti. Praćenjem iskazanog sentimenta studenata kroz vreme mogu se otkriti nagli skokovi u pozitivnom i negativnom sentimentu. Slika 4.46. ilustruje kretanje recenzija prema polaritetu sentimenta i godinama. Vanredna odstupanja u broju recenzija nastavnog osoblja Ekonomskog fakulteta u Subotici zapažena su i u julu 2012. godine, kada je sajt *Oceni profesora* lansiran. U tom periodu je ovaj način evaluacije nastavnog osoblja predstavljao interesantnu novinu za studente, pa se može smatrati i prirodnom pojavom. Zabeležen porast u pozitivnim recenzijama u februaru 2015. godine predstavlja interesantno kretanje i potencijalni je pokazivač da je u naznačenom periodu usledila određena novina u poslovanju, odnosno da je nastavno osoblje izvanredno obavljalo svoj posao. U navedenom periodu značajan je i disbalans između broja pozitivnih i negativnih recenzija. U februaru 2015. godine studenti su pretežno bili zadovoljni dobrom komunikacijom sa profesorima i kvalitetnim predavanjima održanim u zimskom semestru. U novembru 2014. godine primetno je još jedno vanredno odstupanje u pozitivnim komentarima, ne drastično kao u 2015. godini. Studenti su bili zadovoljni profesorima, pretežno bez navođenja specifičnosti svog zadovoljstva. Negativan sentiment je posebno bio izražen u 2012. godini, što se može pripisati lansiranju sajta i naglom i kratkoročnom porastu interesovanja studenata za ovaj vid recenziranja nastavnog osoblja. Globalno posmatrano, broj recenzija sa negativnom semantikom smanjuje se tokom godina, uz vanredna odstupanja u avgustu i novembru 2012. godine, kada studenti pretežno govore o nekorektnosti profesora pri ocenjivanju, juna 2013. godine, kada je vodeći razlog nezadovoljstva studenata nezainteresovanost profesora, i u februaru 2015. godine, kada negativnim komentarima

dominira nezadovoljstvo povezano sa nastavnim materijalima i odnosom profesora prema studentima.



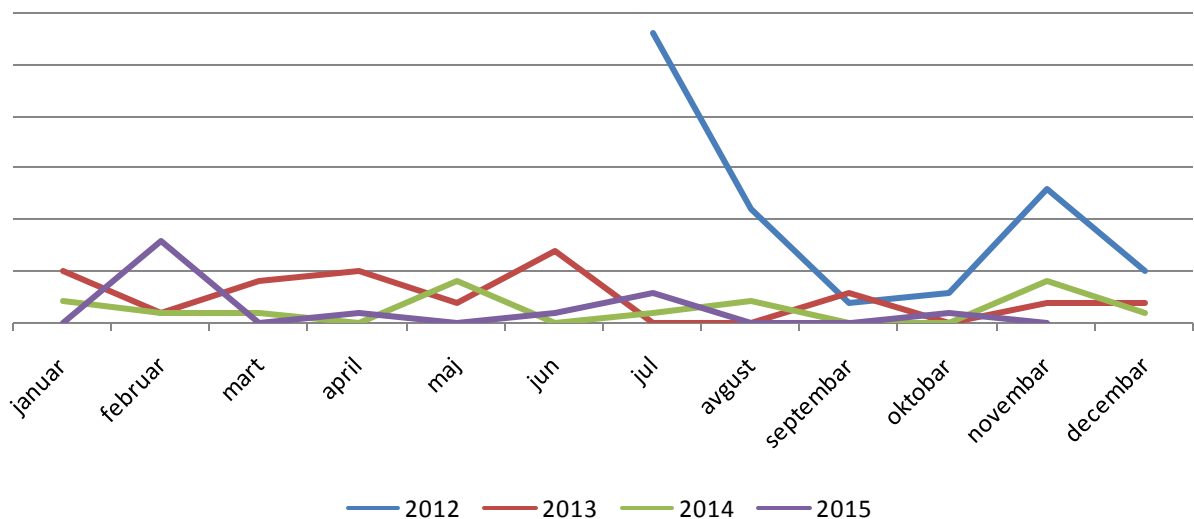
**Slika 4.46. Pregled komentara po polaritetu sentimenta i godinama**

Slika 4.47. ilustruje kretanje pozitivnih recenzija nastavnog osoblja Ekonomskog fakulteta u Subotici kroz period od četiri godine. Dinamika recenzija pozitivne semantike tokom perioda od jula 2012. godine do novembra 2015. godine ne ispoljava ujednačene niti slične pravilnosti. Njihovo kretanje u 2012. i 2013. godini bitno se razlikuje od dinamike u narednim periodima. U 2013. godini prisutni su jasni ciklusi povezani sa ispitnim rokovima, u kojima dolazi do rasta u broju pozitivnih komentara (u junu, septembru i aprilu se održavaju ispitni rokovi za sve studente, dok su mart i novembar meseci u kojima se održavaju apsolventski ispitni rokovi). U 2014. i 2015. godini zabeležen je sličan trend kretanja pozitivnih recenzija u periodu od aprila do avgusta: u maju dolazi do naglog porasta u broju pozitivnih recenzija, čiji pad sledi u junu, a ponovni blaži porast je zabeležen u julu mesecu, u kojem se održava jedan od ključnih ispitnih rokova. Porast zabeležen u maju tokom pomenute dve godine može se povezati sa okončanjem letnjeg semestra, kada studenti imaju jasno formiran stav prema profesorima i motivisani su da ostave recenziju.



Slika 4.47. Kretanje pozitivnih recenzija nastavnog osoblja Ekonomskog fakulteta u Subotici kroz vreme

Kretanje recenzija negativne semantike, prikazano na slici 4.48, karakterišu ciklične oscilacije u svim posmatranim periodima, koje su povezane sa mesecima u kojima se održavaju ispitni rokovi. Pomeranja u zabeleženim ciklusima su posledica promene termina održavanja ispita.



Slika 4.48. Kretanje negativnih recenzija nastavnog osoblja Ekonomskog fakulteta u Subotici kroz vreme

Kretanje pozitivnih recenzija ne karakterišu jasne ciklične oscilacije u svim posmatranim godinama, što može navesti na zaključak da studenti imaju pozitivna iskustva sa profesorima tokom cele godine. Negativna iskustva studenata pretežno su vezana za ispitne rokove ili okončanje semestra, kada studenti mogu da procene upotrebljivost održanih predavanja za

pripremu ispita i kada se stav prema profesoru jasno oformio. U periodima između ispitnih rokova broj negativnih recenzija je ravan nuli ili neznan.

Na osnovu prezentovanih rezultata zaključuje se da adekvatno ekstrahovanje i priprema sadržaja o visokoškolskim institucijama sa sajtova društvenih medija i njihova konsekventna analiza (analiza sentimenta i mišljenja) mogu ukazati na skrivene šablone ponašanja korisnika sajtova. Ovo je ujedno i polazna hipoteza istraživanja, koja se na osnovu opisanih istraživanja i dobijenih rezultata može prihvatiti. Adekvatno ekstrahovanje podataka sa sajta *Oceni profesora* stvorilo je osnovu za kreiranje evaluacionog korpusa nad kojim se uspešno realizovala sentiment analiza razvijenim modelima. Kvalitetna priprema podataka i proces anotacije korpusa osigurali su pouzdanost evaluacionog korpusa. Rezultati primene razvijenih modela, prikazani u ovom poglavlju, ukazuju na široke mogućnosti primene razvijenog pristupa koje se prevashodno odnose na praćenje onlajn reputacije institucije u odnosu na konkurenciju, identifikovanje preferencija postojećih studenata i praćenje preferencija studenata na konkurentskim institucijama, zatim praćenje onlajn reputacije tokom vremena.

Sveobuhvatno praćenje onlajn reputacije institucije zahteva sistematski nadzor i praćenje svih raspoloživih platformi društvenih medija na kojima je institucija prisutna, preuzimanje relevantnih sadržaja i primenu sentiment analize nad njima. Pored sajtova za recenziranje bilo bi korisno u analizu uključiti i sadržaje sa popularnih društvenih mreža (*Facebook* i *Twitter*), sa foruma, kao i komentare na novinske članke. Razvijeni pristup i modeli mogu se primeniti nad novim sadržajima, ali bi proširenje izvora podataka zahtevalo dodatne korake preprocesiranja podataka. Pored koraka preprocesiranja podataka koji su opisani u doktorskoj disertaciji sadržaji prikupljeni sa foruma bi morali da se klasifikuju na činjenice i subjektivne sadržaje koji bi bili predmet sentiment analize. Veliki deo sadržaja na forumima predstavljaju pitanja u kojima korisnici ne iskazuju sentiment, a komunikacija je neretko nefokusirana i odstupa od teme foruma. Iz navedenih razloga, sadržaji sa foruma zahtevaju detaljniju pripremu podataka kako bi se razdvojili subjektivni sadržaji od činjenica, odnosno izdvojili sadržaji koji su u fokusu analize. Komentare na novinske članke uobičajeno karakterišu snažne emocije pisca i pišu se drugačijim stilom u odnosu na recenzije koje su usko fokusirane, sa jasnim stavom i emocijom pisca, i retko kada odstupaju od predmeta recenziranja. Navedene promene u stilu pisanja i vokabularu na različitim sajtovima društvenih medija nameću potrebu za drugačijim sentiment rečnicima, kao i kreiranje reprezentativnih anotiranih korpusa za svaki izvor podataka. Obuhvatanje sadržaja sa drugih izvora društvenih medija i kreiranje neophodnih resursa za primenu postupka sentiment analize izlazi iz okvira doktorske disertacije i može biti predmet budućih istraživanja.

Kvantitativni podaci ukazuju na usko grlo, tj. na problem, dok kvalitativni podaci ukazuju na moguća rešenja identifikovanog problema. Uvid u probleme, kao i smernice za njihovo rešavanje nisu automatizovani, te koraci koji će se preduzeti umnogome zavise od onoga što se otkrilo sentiment analizom. Rezultati sentiment analize nisu dovoljni za kreiranje efikasnih poslovnih strategija. Potrebno ih je integrisati sa saznanjima do kojih se dolazi primenom poslovne inteligencije, sa znanjima ljudi eksperata i važnim metrikama poslovanja. Dakle, sentiment kao

samostalna metrika nije dovoljna, ali predstavlja prekretnicu za oblikovanje budućih analiza i smernicu za unapređenje poslovanja.



## 5. ZAKLJUČAK

Sa pojavom sajtova društvenih mreža i medija, kao interaktivnih platformi, sve veći deo diskusije javnosti i kreiranja potrošačkih odluka odvija se na webu. Recenzije potrošača i korisnika usluga predstavljaju istaknut primer korisnički generisanog sadržaja u kojima ljudi iskazuju svoja mišljenja i iskustva na webu. Ovakav glas potrošača, kao vredan izvor informacija za pojedince i kompanije, ukazuje na potrebu za razvojem alata koji će automatski analizirati i sumirati navedene tekstualne podatke. Doktorskom disertacijom se nastojalo ukazati na značaj prikupljanja korisnički generisanog sadržaja sa sajtova društvenih mreža i društvenih medija, kao i na njihovu konsekvantnu analizu.

Istraživanje koje je predmet doktorske disertacije objedinjuje različite istraživačke oblasti. U uvodnom poglavlju dat je pregled i opis osnovnih koncepata i istraživačkih oblasti obuhvaćenih doktorskom disertacijom: oblast društvenih medija i društvenih mreža i oblast analize podataka (*data mining* i sentiment analiza), kao i predmetna terminologija koja se koristi u radu.

Na našem govornom području sentiment analiza predstavlja neistraženu oblast i retko korišćen pristup u poslovanju. Iz ovog razloga je u drugom poglavlju dat pregled dosadašnjih istraživanja iz oblasti sentiment analize, rađenih pretežno za englesko govorno područje, čime je ukazano na šire mogućnosti ovog analitičkog pristupa. Takođe, sagledani su rezultati dosadašnjih istraživanja o primeni sentiment analize u visokom obrazovanju.

Treće poglavlje daje detaljniji prikaz metoda, tehnika i pristupa koji se primenjuju u sentiment analizi, kao i izazove sa kojima su dosadašnja istraživanja bila suočena. Izučavanjem postignuća u oblasti sentiment analize, kao i razvijenih metoda, tehnika i pristupa omogućeno je sticanje uvida o njihovoj primenljivosti na oblast visokog obrazovanja, čime se odogovorilo na sledeći istraživački cilj:

*Utvrđiti primenljivost metoda, tehnika i pristupa analize sadržaja sa društvenih medija na oblast visokog obrazovanja, kao i moguće efekte koje bi njihova primena dala na usluge i servise koje visokoškolske institucije nude.*

Usled nedostatka razvijenih resursa, konstatovana je neophodnost kreiranja evaluacionog korpusa koji će omogućiti primenu algoritama nadgledanog učenja u sentiment analizi. Adekvatno kreiran korpus, koji podrazumeva manuelnu anotaciju skupa tekstualnih podataka, otvara i mogućnost primene klasifikacije sentimenta na osnovu leksikona koji se deriviraju iz korpusa. Takođe, anotiranjem negacije u fazi razvoja korpusa omogućeno je njeno uvažavanje pri identifikovanju sentiment polariteta, kao i razvoj budućih sistema za detektovanje negacije u sadržajima. Atributi za realizaciju sentiment analize mogu se definisati na osnovu prisustva i frekventnosti reči u korpusu. Upotreba vrste reči kao atributa za sentiment analizu, dobijenih POS označavanjem, u našem slučaju nije bila moguća, jer ne postoje javno dostupni alati za realizaciju ovog zadatka u tekstovima na srpskom jeziku. Razvoj alata za POS označavanje

zalazi u domen lingvistike i nije predmet doktorske disertacije. Dosadašnja istaživanja ukazuju na sledeće efekte primene sentiment analize u domenu visokog obrazovanja: unapređenje inteligentnih tutorskih sistema, identifikovanje prednosti i nedostataka predmeta, identifikovanje problema sa kojima se studenti suočavaju i njihovo blagovremeno rešavanje tokom trajanja semestra, predviđanje studenata koji će napustiti kurs koji pohađaju, odnosno fakultet, i podsticanje angažovanja studenata na nastavi. Iako se u drugim domenima korisničke recenzije intenzivno koriste za sveobuhvatno praćenje onlajn reputacije i zadovoljstva korisnika, prema našem saznanju, nema istraživanja koja se bave pitanjem studentskog zadovoljstva i praćenjem onlajn reputacije visokoškolske institucije analizom studentskih recenzija nastavnog osoblja, što predstavlja predmet istraživanja doktorske disertacije.

Centralni deo doktorske disertacije predstavlja istraživanje prikazano u četvrtom poglavlju, koje daje odgovor na sledeći istraživački cilj:

*Izraditi metodološki okvir za ekstrahovanje sadržaja o pominjanjima visokoškolskih institucija Republike Srbije na društvenim medijima i sprovođenje postupka sentiment analize za potrebe upravljanja reputacijom.*

U ovom poglavlju detaljno je opisan proces prikupljanja korisnički generisanog sadržaja sa portala *Oceni profesora*, neophodno preprocesiranje podataka koje prethodi kreiranju korpusa za domen visokog obrazovanja postupkom anotacije prikupljenog sadržaja i analiza kreiranog korpusa primenom razvijenih sentiment rečnika i algoritama mašinskog učenja. Navedeni koraci predstavljaju sveobuhvatan metodološki okvir koji omogućava sprovođenje postupka sentiment analize primenom nadgledanog učenja i leksikona.

Ograničenja sentiment analize leže u činjenici da je mašinsko učenje još uvek u ranoj fazi razvoja i da se emocije teško mogu računarskim putem predvideti usled kulturoloških faktora, nijansi u jeziku i stilu pisanja, kao i različitog konteksta u kojem se pojedine reči ili fraze koriste, što značajno otežava transformisanje niza pisanog teksta u jednostavni sentiment koji će iskazivati saglasnost ili nesaglasnost. Takođe, alati za sentiment analizu imaju poteškoću da identifikuju ironiju, humor i suptilnosti u ljudskom govoru, poput efekta koji emotikoni imaju na značenje i nameru napisanih grubih ili negativnih reči. Navedeno je nametnulo sledeći istraživački cilj:

*Identifikovati ograničenja i poteškoće u analizi podataka koje se mogu pojaviti kao posledica specifičnog predmeta analize i usled specifičnosti jezika i ukazati na potencijalna rešenja.*

Realizacijom polaznih koraka istraživanja, opisanih u [Poglavlju 4.1.1](#), identifikovana su sledeća ograničenja i poteškoće:

1. Neformalni stil pisanja, bez poštovanja pravopisnih pravila predstavlja jednu od osnovnih karakteristika korisnički generisanog sadržaja na vebu. U većini recenzija, koje su

prikupljene sa sajta *Oceni profesora*, identifikovan je izostanak praznog mesta nakon interpunkcijskih znakova, upotreba višestrukih uzvičnika i tačaka na polovini ili na kraju rečenice, izostanak interpunkcijskog znaka na kraju rečenice, dijakritički znakovi se nisu koristili i načinjeni su brojni gramatički propusti. Za korektno sprovođenje sentiment analize neophodno je izvršiti korekciju pravopisnih grešaka. Trenutno postoji nekoliko alata koji imaju ugrađen rečnik srpskog jezika i mogli bi se koristiti za korigovanje pravopisnih grešaka. Međutim, u praksi ovi alati ne rade ispravno. Aspell i Hunspell alati, koji su inicijalno korišćeni za ove potrebe, rade sa brojnim propustima i greškama. Aspell alat ne prepoznaje reči sa dijaktičkim znacima. Hunspell alat nema mogućnost poništavanja prethodne akcije čime je onemogućeno korigovanje grešaka. Dodatno, oba alata zahtevaju korigovanje jedne po jedne reči u prikupljenim sadržajima. Iz ovih razloga upotrebljen je onlajn alat Hascheck koji je namenjen korigovanju pravopisnih grešaka u sadržajima pisanim na hrvatskom jeziku. Navedenim alatom mogu se korigovati pojedinačna pojavljivanja reči u skupu, kao i korekcija reči na svim mestima pojavljivanja, čime se štedi vreme. Dodatno, daje mogućnost unošenja ispravnog oblika reči ukoliko softver u svojoj bazi podataka nema adekvatan oblik. Ograničenje ovog alata se ogleda u maloj količini podataka koju može da procesuiru u jednoj analizi. Ovakva situacija navodi na zaključak da za srpski jezik još uvek nije razvijen adekvatan alat za korigovanje pravopisnih grešaka ili nije javno dostupan.

2. U sentiment analizi osnovna priprema podataka podrazumeva segmentaciju teksta: razdvajanje sadržaja na rečenice (engl. *sentence splitting*) i razdvajanje rečenica na reči, odnosno tokenizaciju reči. Upotreba interpunkcijskih znakova, prevashodno tačke, predstavlja osnovu rada razvijenih alata za podelu sadržaja na rečenice. Ovakvi alati su razvijeni za dominantne svetske jezike, poput engleskog, nemačkog i drugih, i uključuju specifičnosti ovih jezika, koje se prvenstveno ogledaju u upotrebi tačke u skraćenicama nakon kojih ne sledi kraj rečenice. U srpskom jeziku koristi se set skraćenica karakterističan isključivo za naše govorno područje. Takođe, datumi i redni brojevi se pišu drugačijim navođenjem. Navedene specifičnosti srpskog jezika onemogućile su ispravnu primenu već razvijenih alata za segmentaciju sadržaja na rečenice nad sadržajima pisanim na srpskom jeziku. Za potrebe istraživanja koje je predmet doktorske disertacije razvijen je alat za razdvajanje sadržaja na rečenice, koji obuhvata istaknute specifičnosti, a ima mogućnost i dodatnog proširenja. Algoritam koji predstavlja osnovu rada razvijenog alata prikazan je u [Prilogu B](#).
3. Lematizacija i tzv. *stemming* takođe predstavljaju uobičajene korake pripreme podataka. Cilj oba postupka jeste uklanjanje fleksija, odnosno deriviranih oblika reči, čime se svaka reč svodi na svoj osnovni oblik – lemu. Množina u srpskom jeziku je primer fleksije: „profesor“ i „profesori“. Prema našem saznanju, alati za lematizaciju sadržaja ne postoje ili nisu javno dostupni za srpski jezik. Iz ovog razloga i usled činjenice da je razvoj ovakvog alata usko povezan sa lingvistikom, navedeni korak preprocesiranja sadržaja nije realizovan i biće predmet budućih istraživanja. Očekivano je da bi lematizacija

sadržaja podigla preciznost rada razvijenih modela, jer bi se smanjila kompleksnost pisanog sadržaja svođenjem svake reči na njen osnovni oblik, čime bi postojalo manje varijacija u analiziranim sadržajima.

Naredni istraživački cilj na koji doktorska disertacija daje odgovor, a koji predstavlja centralni deo predloženog metodološkog okvira za primenu sentiment analize, jeste:

*Formirati anotirani korpus kao osnovu za sprovođenje sentiment analize pominjanja visokoškolskih institucija.*

Kreiranje sopstvenog evaluacionog korpusa motivisano je, pre svega, nedostatkom već razvijenih korpusa za naše govorno područje. Takođe, usled razlika u domenima primene sentiment analize i tipovima tekstova nad kojima se primenjuje (recenzija, poruka na forumu, komentar na novinski članak i drugo), ne postoji standardizovana shema anotacije koja bi poslužila kao osnova za kreiranje referentnog korpusa za specifičan zadatak sentiment analize. Buduća istraživanja bi trebalo da se bave definisanjem opšte sheme anotacije za najznačajnije zadatke sentiment analize i tipove tekstova, odnosno domene. Na ovaj način bi se stvorila osnova za kvalitetnu komparaciju istraživačkih pristupa. Postupak anotacije i eksploracija kreiranog korpusa opisane su u [Poglavlju 4.1.3](#). Da bi se shvatio problem i obezbedila osnova za primenu sentiment analize nad prikupljenim sadržajima, bilo je neophodno osmisлити formalni model koji opisuje elemente iskazanog sentimenta u tekstovima. Iz nestruktuiranih podataka izvedena je hijerarhija aspekata, odnosno tema o kojima studenti pišu u recenzijama, sa tri aspekta na višem hijerarhijskom nivou (*Profesor, Predmet i Drugi aspekt*) i četiri aspekta na nižim nivoima (*Predavanja i Odnos prema studentima, odnosno Materijali i Organizacija predmeta*). U narednom koraku analiziran je način na koji se sentiment iskazuje prema idenentifikovanim aspektima. Ustanovljeno je da studenti u recenzijama iskazuju tri tipa sentimenta: pozitivan, negativan i neutralan, kao i da je prisutan različit intenzitet pozitivnog i negativnog sentimenta. Otuda, model anotacije prikupljenih podataka na nivou rečenice podrazumeva anotaciju aspekta, iskazanog sentimenta i njegovog intenziteta, dok model anotacije na nivou reči podrazumeva označavanje individualnih sentiment izraza, negacije i njenog opsega. Na osnovu razvijenih modela kreirana je anotaciona shema prikazana u [Prilogu C](#), koja je predstavljala osnovu za manuelno označavanje studentskih recenzija. Detaljno uputstvo za anotaciju, prikazano u [Prilogu D](#), može poslužiti za buduće proširenje ili adaptaciju razvijenog korpusa. Rezultat anotacije predstavlja korpus, koji je omogućio detaljnu analizu problema, analizu lingvističkih fenomena i stvorio osnovu za procenu relevantnosti pojedinih potproblema. Za evaluaciju kvaliteta anotacije upotrebljeni su *Fleiss kappa* i *agr* metrika. Postignuta saglasnost anotatora ukazuje na konzistentnost u anotaciji i visoku pouzdanost razvijenog korpusa za domen visokog obrazovanja. Rezultati procene stepena saglasnosti anotatora prikazani su u [Poglavlju 4.1.3.2](#). Analizom korpusa, prikazanom u [Poglavlju 4.1.3.3](#), izvedeni su sledeći zaključci i rezultati:

1. Korpus je sačinjen iz 3.863 recenzija i 6.896 rečenica napisanih na srpskom jeziku, koje su preuzete sa sajta *Oceni profesora*.

2. Recenzije studenata su usko fokusirane. U 98.22% recenzija predmet recenziranja predstavlja domen visokog obrazovanja, dok je u 98.7% recenzija iskazan sentiment prema nekom od predefinisanih aspekata.
3. Po obimu, većina recenzija je srednje dužine (od 6 do 30 reči), što zahteva identifikovanje šireg spektra veoma specifičnih karakteristika za realizaciju klasifikacije.
4. Distribucija recenzija prema polaritetu i aspektima je asimetrična. Korpusom dominiraju pozitivni komentari i aspekt *Profesor*. Negativne recenzije su po obimu duže od pozitivnih i upotrebljen vokabular je raznovrsniji, što potencijalno može uzrokovati poteškoće u postizanju dobrog *Recall*-a za klasu negativnih recenzija. Neutralan sentiment se retko pojavljuje u korpusu i ova klasa sentimenta predstavlja poseban izazov, kao i u većini ranijih istraživanja. Pošto je u većini recenzija iskazan negativan ili pozitivan sentiment, neutralna kategorija se ne mora uključiti u sistem za automatsko prepoznavanje sentimenta.
5. Korpusom dominiraju sentiment izrazi snažnog intenziteta, čineći 65% identifikovanog sentimenta.
6. Tokom anotacije korpusa identifikovane su pozitivne i negativne sentiment reči, prikazane u [Prilogu F](#). Postojeći sentiment rečnici se nisu mogli koristiti u analizi sentimenta zbog domenske zavisnosti prikupljenog sadržaja i specifičnog stila pisanja. Unutar domena uočena je kontekstualna zavisnost pojedinih reči ili fraza. Na primer, zavisno od konteksta u kojem se koristi, reč *blago* menja značenje i sentiment polaritet sadržaja. Takođe, identifikovan je set fraza, prikazan u [Prilogu H](#), koje studenti često koriste u recenzijama, a koje nose jasan pozitivan ili negativan sentiment. Na osnovu identifikovanih sentiment reči i fraza u korpusu kreirani su sentiment rečnici. Postupak njihovog kreiranja i forma rečnika prikazana je u [Poglavlju 4.1.4](#). Razvijeni rečnici predstavljaju osnovu za sprovođenje klasifikacije sentimenta na osnovu leksikona.
7. Slično drugim jezicima za koje su razvijeni alati za automatsko prepoznavanje sentimenta, u srpskom jeziku sentiment reči su pretežno pridevi. Navedeno saznanje navodi na zaključak da bi POS označavanje sadržaja bilo korisno za automatsko izdvajanje sentiment rečnika.
8. Izdvojeni su intenzifikatori i neutralizatori sentimenta, prikazani u [Prilogu G](#). Reč je o specifičnom setu reči koje se upotrebljavaju u domenu visokog obrazovanja, a koje, zavisno od konteksta u kom se koriste, mogu pojačati ili umanjiti iskazani sentiment u sadržaju. U funkciji intenzifikatora sentimenta javljaju se prilozi (za količinu i način), pridevi, rečce i veznici. Na osnovu izdvojenih reči i fraza kreirani su rečnici intenzifikatora i neutralizatora koji su upotrebljeni u kasnijim analizama sadržaja.
9. Stil pisanja u recenzijama omogućava jednostavno detektovanje negacije usled nedvosmislene upotrebe ključnih reči negacije. Međutim, problem donosi pojava višestruke negacije, koja sa jedne strane može uzrokovati promenu sentiment polariteta i značenja rečenice, dok su u drugačijem kontekstu ovakve promene izuzete – sentiment polaritet uprkos pojavi negacije ostaje nepromenjen. Otuda, sentiment polaritet nije

isključivo pod uticajem negacije koliko je pod uticajem imenice, prideva ili glagola koji se negira. Identifikovani oblici negacije prikazani su u [Prilogu I](#). Na osnovu njih je kreiran rečnik negacije upotrebljen u klasifikaciji sentimenta na osnovu leksikona.

10. Zavisne uslovne rečenice se koriste za iskazivanje pozitivnih komentara izrazitog intenziteta.
11. Modalni glagoli (npr. moći, trebati) i komparativi prideva i priloga (npr. bolje) koriste se za iskazivanje negativnog sentimenta. Takođe, modalni glagoli u potencijalu (npr. trebalo bi, moglo bi) koriste se za iskazivanje preporuka u komentarima.

Iskazivanje mišljenja u recenzijama predstavlja kompleksan lingvistički fenomen. Mišljenja se iskazuju eksplicitno i implicitno. Pojedine reči jačaju intenzitet iskazanog sentimenta, slabe ga ili menjaju polaritet sentimenta. U pojedinim rečenicama prisutni su višestruki iskazi sentimenta usmereni ka različitim aspektima i često su iskazani sentiment različitog polariteta. Rezultati pokazuju da su ovakve kompleksne konstrukcije ređe prisutne u korpusu. Na primer, manje od 2% iskazanog sentimenta je neutralno. Međutim, više od 20% skupa za obučavanje čine recenzije u kojima je iskazan višestruki sentiment različitog polariteta. Identifikovane specifičnosti i suptilnosti u stilu pisanja i načinu na koji se iskazuje sentiment izuzetno su bitne za adekvatno lingvističko modelovanje iskazanog mišljenja, ali u dizajniranju sistema za automatizovano prepoznavanje sentimenta potrebno je akcenat staviti na prepoznavanje najčešćih paterna. Otuda se većina sentiment iskaza u studentskim recenzijama može identifikovati primenom jednostavnijih pristupa, poput primene odgovarajućih sentiment rečnika i pristupa najbližih suseda. Pristup zasnovan na leksikonima predstavlja dobru polaznu tačku u sentiment analizi prikupljenih sadržaja. Prednost ovog pristupa se ogleda i u mogućnosti proširenja, kao što bi bila integracija sentiment rečnika i sistema zasnovanih na pravilima. Budući pravci istraživanja ogledaće se u analizi mogućnosti automatskog generisanja sentiment rečnika, a izazov predstavljaju reči koje se retko pojavljuju u korpusu te je moguće da ih sistem ne prepoznaje kao relevantne, što pri manualnoj anotaciji sentiment reči nije slučaj.

Anotacije sadržaja na nivou rečenice i reči, date za svaku recenziju, rezultovale su neophodnim informacijama za realizaciju klasifikacije recenzija prema sentiment polaritetu ili različitim aspektima, za identifikovanje aspekata sa najsnažnijim i najslabijim intenzitetom sentimenta, identifikovanje karakteristika koje studenti preferiraju u svakom aspektu, komparaciju sentimenta studenata sa različitim visokoškolskih institucija, kao i monitoring i analizu promena u mišljenjima i iskazanom sentimentu tokom vremena. Sve ovo je omogućilo da se steknu vredna saznanja o stavovima i mišljenju studenata o visokoškolskoj instituciji koju pohađaju. Razvijeni modeli koriste samo određene identifikovane specifičnosti o stilu pisanja i strukturi rečenica, koje se odnose na intenzifikatore i neutralizatore sentimenta i negaciju. Međutim, detaljna zapažanja, dokumentovana u [Poglavlju 4.1.3.3](#), stvorila su osnovu za generisanje specifičnih pravila rada automatizovanih sistema za prepoznavanje sentimenta, koja će biti predmet budućih istraživanja.

U nastavku teze, u [Poglavljima 4.1.5. i 4.1.6.](#), prikazani su razvijeni modeli sentiment analize i evaluacija njihovih performansi. Sprovedeni eksperimenti omogućili su *identifikovanje najadekvatnijeg načina za sentiment analizu srpskog jezika, odnosno sentiment analizu pominjanja visokoškolskih institucija, kroz sistematičnu analizu i poređenje odabranih algoritama*, što predstavlja naredni istraživački cilj.

Problem identifikovanja sentimenta u studentskim recenzijama definisan je kao binarni problem klasifikacije (pozitivan sentiment nasuprot negativnom) u modelima baziranim na primeni algoritama mašinskog učenja ili kao trostruki problem klasifikacije (pozitivan, negativan i neutralan sentiment) u modelima baziranim na primeni sentiment leksikona. Identifikovanje seta atributa, koji će se koristiti u klasifikacionim modelima, predstavlja važan korak. Otuda je realizovan niz eksperimenata kako bi se identifikovao optimalan set atributa. U nastavku su sistematizovani osnovni rezultati preduzetih eksperimenata:

1. U klasifikaciji sentimenta primenom algoritama mašinskog učenja upotreba bigrama kao atributa klasifikacije daje superiornije rezultate u poređenju sa unigramima, trigramima i četirigramima.
2. Filtriranje stop reči daje značajan pomak u tačnosti klasifikacije.
3. Najbolje performanse u klasifikaciji sentimenta primenom algoritama mašinskog učenja na nivou recenzija postižu se primenom SVM algoritma – 84.94% tačnosti i F-mera iznosi 89.66% za pozitivnu klasu, a 72.32% za negativnu. kNN algoritam, gde je  $k=3$ , rezultuje nešto nižom tačnosti od 84.53%, ali približno uspešno klasifikuje recenzije u skupu za obučavanje i u skupu za testiranje i rezultuje nešto višom F-merom za negativnu klasu – 74.28%. U kontekstu poslovanja relevantnija je negativna klasa koja ukazuje na nezadovoljstvo studenata i potrebna poboljšanja u poslovanju. Iz tog razloga zaključuje se da je kNN algoritam bolje rešenje za dati domen.
4. U klasifikaciji sentimenta primenom algoritama mašinskog učenja na nivou rečenice najbolje performanse daje SVM algoritam – 80.13% tačnosti u klasifikovanju nepoznatih primera rečenica. F-mera za pozitivnu klasu iznosi 84.53%, dok za negativnu iznosi 72.21%.
5. U klasifikaciji sentimenta primenom sentiment leksikona najbolji rezultati se postižu uvođenjem zasebnih rečnika za pozitivne i negativne sentiment reči, intenzifikatore, neutralizatore, i negaciju i druge reči koje menjaju sentiment naredne reči u rečenici.
6. Manuelno korigovanje rečnika značajno poboljšava rezultate klasifikacije.
7. Tačnost klasifikacije na nivou recenzija iznosi 80.71%. F-mera iznosi 89.75% za pozitivnu klasu, 77.28% za negativnu i 26.04% za neutralnu.
8. Tačnost klasifikacije na nivou rečenica iznosi 73.70%. F-mera iznosi 86.05% za pozitivnu klasu, 71.80% za negativnu i 4.49% za neutralnu.
9. Identifikovanje neutralne klase predstavlja najveći izazov, što je potvrđeno i u drugim istraživanjima ((Khan, et al., 2014), (Kasper, et al., 2011)). Nezastupljenost neutralne

klase u korpusu može se smatrati primarnim uzrokom niske preciznosti njene klasifikacije.

10. Na nivou recenzija bolje performanse se postižu primenom klasifikacije zasnovane na leksikonima, dok se na nivou rečenica bolje performanse postižu primenom algoritama mašinskog učenja.
11. Kategorija lažno negativnih vrednosti predstavlja problem podjednako kod pristupa zasnovanom na algoritmima mašinskog učenja, kao i kod pristupa zasnovanom na sentiment leksikonima. Reč je o situacijama kada sistem identifikuje prisustvo određene negativne reči u analiziranom sadržaju, ali ne prepozna sveukupno pozitivan kontekst.

Iako razvijeni modeli sentiment analize daju dobre rezultate, ostaje otvoreno pitanje daljih mogućnosti unapređenja kroz lematizaciju sadržaja, anotaciju sentiment polariteta na nivou celokupne recenzije, anotaciju neutralnih sentiment izraza radi obogaćivanja sentiment rečnika, čime bi se pospešila preciznost rada modela i, eventualno, tačnost klasifikacije.

Poslednji cilj doktorske disertacije odnosi se na *sagledavanje mogućih načina unapređenja poslovanja visokoškolskih institucija oslanjanjem na rezultate analiza sadržaja društvenih medija*. Mogućnosti upotrebe razvijenih modela za praćenje onlajn reputacije i identifikovanje izvora nezadovoljstva studenata ilustrovane su na primeru Ekonomskog fakulteta u Subotici i detaljno opisane u [Poglavlju 4.2](#). Rezultati primene klasifikacije sentimenta sumirani su i vizualizovani prema različitim kriterijumima.

Razvijeni korpus i realizovana sentiment analiza stvorili su uslove za:

1. Identifikovanje globalnog zadovoljstva studenata sumiranjem pozitivnih i negativnih recenzija na nivou institucije.
2. Komparaciju onlajn reputacije institucija poređenjem procenta pozitivnih i negativnih recenzija po institucijama.
3. Analizu pozitivnih i negativnih recenzija po aspektima kako bi se identifikovao aspekt poslovanja visokoškolske institucije koji studenti favorizuju, odnosno aspekt koji u najvećoj meri kritikuju.
4. Komparaciju institucija prema iskazanom zadovoljstvu i nezadovoljstvu po aspektima.
5. Analizu asocijativnih pravila, bigrama i trigrama unutar pozitivnih i negativnih recenzija, čime se stiče uvid u dominantne i važne fraze kojima studenti iskazuju svoje zadovoljstvo ili nezadovoljstvo, a koje ukazuju na područja poslovanja koja daju komparativnu prednost, odnosno koja treba unaprediti, respektivno.
6. Praćenje iskazanog sentimenta studenata kroz vreme, sa ciljem identifikovanja oscilacija u zadovoljstvu i nezadovoljstvu studenata.

Na osnovu sprovedenih analiza zaključeno je da su studenti Ekonomskog fakulteta u Subotici pretežno zadovoljni iskustvom studiranja.



U poređenju sa konkurentskim fakultetima, prema komentarima ostavljenim na sajtu *Oceni profesora*, zaključuje se da je Ekonomski fakultet u Subotici bolje ocenjen od strane svojih studenata po pitanju *Odnosa profesora prema studentima*, te da je prednost Ekonomskog fakulteta u Subotici kvalitetan odnos nastavnika prema studentima i postojanje kvalitetnih nastavnih materijala i udžbenika. Korpusom dominira aspekt *Profesor*, koji se u većoj meri ocenjuje kao pozitivan.

Međutim, u poređenju sa drugim relevantnim visokoškolskim institucijama, studenti Ekonomskog fakulteta u Subotici iskazuju viši stepen nezadovoljstva po pitanju *Predavanja* i *Predmeta*, što se ogleda u lošijim komentarima. Kao najveći izvor nezadovoljstva studenata identifikovan je aspekt *Predavanja*. Studenti oštrim rečnikom i intenzivnim sentimentom kritikuju kvalitet pojedinih predavanja, opisujući ih kao neinteresantna i zastarela. Nadalje, iz komentara se vidi da profesori nedovoljno uključuju studente u nastavni proces, što utiče na (ne)zainteresovanost studenata i (ne)pažnju na času. Studenti oštro kritikuju i nezainteresovanost pojedinih profesora za predavanja, didaktičke metode, tj. način na koji se znanje prenosi, i forsiranje učenja gradiva napamet i „od reči do reči“.

Praćenjem kretanja sentiment polariteta u recenzijama kroz vreme identifikovane su jasne ciklične oscilacije u negativnim komentarima, koje su povezane sa ispitnim rokovima, dok u pozitivnim komentarima slični šabloni kretanja nisu zapaženi.

Realizacijom postavljenih istraživačkih ciljeva i sprovedenim eksperimentima, koji su opisani u prethodnim poglavljima doktorske teze, pokazano je da *adekvatno ekstrahovanje i priprema sadržaja o visokoškolskim institucijama sa sajtova društvenih medija i njihova konsekvantna analiza (analiza sentimenta i mišljenja) ukazuju na skrivene šablone ponašanja korisnika sajtova*. Osluškivanjem sentimenta studenata u komentarima o instituciji na društvenim mrežama i medijima, visokoškolskim institucijama se pruža mogućnost deriviranja bitnih saznanja i prevođenja stečenih saznanja u akcije, te se sa jedne strane mogu preduzeti korektivne mere kako bi se prevazišli identifikovani problemi, a sa druge strane, otvara se mogućnost upotrebe stečenih saznanja za preduzimanje mera ka unapređenju studentskog iskustva i sveukupne slike javnosti o instituciji. Primena razvijenih modela sentiment analize stvorila je uslove za praćenje onlajn reputacije institucija, za komparaciju sa drugim institucijama, identifikovanje preferencija i nezadovoljstva studenata i ukazala na široke mogućnosti primene sentiment analize u domenu visokog obrazovanja.

Rezultati istraživanja doprinose boljem razumevanju kompleksnosti sentiment analize. Teorijski doprinos doktorske disertacije ogleda se u sistematizaciji relevantnih teorijskih koncepata koji objašnjavaju aktuelna dostignuća i mogućnosti u implementaciji sentiment analize. Usled činjenice da je primena sentiment analize nezastupljena istraživačka oblast u Srbiji, sistematizacijom vladajućih stavova, metoda, tehnika i pristupa kreirana je adekvatna biblioteka jedinica na srpskom jeziku. Naučno-istraživački doprinosi doktorske disertacije odnose se na:

1. Sistematizaciju metodoloških koraka i formiranje metodološkog okvira za primenu sentiment analize;
2. Razvoj alata za preprocesiranje sadržaja na srpskom jeziku, tzv. razdelnika rečenica;
3. Razvoj evaluacionog korpusa na srpskom jeziku za domen visokog obrazovanja, čime su stvoreni neophodni preduslovi za implementaciju različitih pristupa sentiment analizi i njihovu evaluaciju. Javno dostupni korpusi na srpskom jeziku potpuno izostaju, te ovo predstavlja jedinstven doprinos, prvi ove vrste;
4. Izradu uputstva za anotaciju korpusa, koje može poslužiti za buduće proširenje i adaptaciju korpusa;
5. Kreiranje sentiment rečnika, rečnika intenzifikatora i neutralizatora, kao i rečnika reči koje menjaju značenje i sentiment naredne reči u rečenici;
6. Detaljnu lingvističku analizu sadržaja, koja će poslužiti za generisanje specifičnih pravila na osnovu kojih će raditi budući sistemi za automatsko prepoznavanje sentimenta;
7. Razvoj modela i alata za sentiment analizu sadržaja na srpskom jeziku;
8. Sistematizaciju mogućnosti primene razvijenih modela u praćenju onlajn reputacije i za unapređenje poslovanja visokoškolskih institucija, kao i postavljanje smernica za korektivne akcije.

Kao većina doktorskih disertacija, i ova teza inicira niz istraživačkih tema koja će biti predmet budućih istraživanja. Na ovom mestu ćemo pomenuti tri anticipirana pravca istraživanja.

#### 1. Sentiment analiza sadržaja sa različitih izvora podataka

Fokus istraživanja doktorske teze predstavlja sentiment analiza studentskih recenzija. Prezentovani modeli i pristupi su razvijani ka ovom izvoru korisnički generisanih sadržaja. Međutim, povratne informacije studenata o poslovanju visokoškolskih institucija prisutne su i u drugim formatima. Studenti iskazuju svoja mišljenja na mikroblogovima, forumima i društvenim mrežama, poput *Facebook*-a. Pristup sentiment analizi primenjen u istraživanju potrebno je primeniti i nad sadržajima sa navedenih izvora podataka radi sticanja sveobuhvatne slike o onlajn reputaciji visokoškolske institucije. Međutim, istraživanje predstavljeno u doktorskoj disertaciji je slično drugim istraživanjima iz domena sentiment analize: uzima u obzir jedan, uzan, domen aplikacije. Iz ovoga se otvara pitanje, koje će biti predmet budućih istraživanja, o primeni razvijenih pristupa i modela nad sadržajima iz drugih izvora.

#### 2. Proširivost rezultata na druge domene

U prilagođavanju ili primeni naših pristupa u drugim domenima prvenstveno treba voditi računa o različitim tipovima tekstova, koji nose razlike u stilovima pisanja i dužini tekstova (recenzije studenata na sajtu *Oceni profesora* po dužini su slične sadržajima koje se objavljuju na *Twitter*-u, dok komentari na novinske članke ili objave na forumima mogu biti duže). Iz ugla sentiment analize važno je upoznati se sa karakteristikama različitih tipova tekstova. Buduća istraživanja opisace kvalitativno i kvantitativno ove razlike. Ovakva analiza bi omogućila sistematično

ispitivanje mogućnosti adaptacije razvijenih pristupa i modela na različite domene. Otvoreno pitanje je i da li se modeli obučeni za klasifikaciju iskazanog sentimenta u recenzijama mogu direktno upotrebiti za klasifikaciju sentiment polariteta u blogovima, mikroblogovima ili komentarima na novinske članke. Takođe, povlači se pitanje i dostupnosti ovakvih podataka.

### 3. Detekcija sentimenta uticajnih korisnika

Uz sentiment analizu potrebno je pratiti i analizirati uticajne korisnike na društvenim mrežama i medijima. Portal *Oceni profesora* ne daje mogućnost praćenja uticajnih korisnika, jer se sve recenzije ostavljaju anonimno. Drugi sajtovi društvenih mreža i medija zahtevaju da korisnik ostavlja komentar ili recenziju uz svoje korisničko ime. Praćenje uticajnih korisnika podrazumeva identifikovanje njihovih negativnih komentara, praćenje dometa koji njihovi sadržaji imaju i praćenje broja ljudi na koje će sadržaj imati uticaja. Na ovaj način je omogućeno blagovremeno neutralizovanje pominjanja koja u najvećoj meri mogu narušiti reputaciju institucije. Takođe, predmet budućih istraživanja biće i povezivanje sentiment izraza i aspekata i generisanje modela koji će uspešno identifikovati predmet pisanja u recenzijama.

Sistemi sentiment analize i derivirani rezultati moraju se uvek kombinovati sa ljudskom ekspertizom. Za razliku od ljudi, automatizovani sistemi nemaju mogućnost prepoznavanja suptilnosti, poput govora tela i sarkazma, koje doprinose značenju i kontekstu. Otuda ljudi obezbeđuju kontekst za pojedine reči ili fraze i ukazuju na aspekte kojima je potrebno obučiti sistem, angažovani su u postavljanju kategorija prema kojima će se komentari klasifikovati, definisanju poslovnih ciljeva i interpretaciji rezultata sentiment analize. Međutim, jaz između tačnosti mašinske analize i analize koju realizuju ljudi drastično se smanjuje usled razvoja alata i tehnologije sentiment analize.

Urađena doktorska disertacija ima širi društveni doprinos. Rezultate istraživanja mogu koristiti sve obrazovne institucije i univerziteti koji su zainteresovani za unapređenje odnosa sa studentima i praćenje onlajn reputacije. Usled domenske zavisnosti, određene modifikacije se zahtevaju u primeni rezultata istraživanja u poslovanju kompanija, dok se sam pristup sentiment analizi, opisan u doktorskoj disertaciji, može upotrebiti bez ograničenja u poslovanju kompanija iz bilo koje privredne delatnosti. Institucije i kompanije koje prve u potpunosti inkorporiraju sentiment analizu sa drugim alatima i metrikama, steći će značajnu poslovnu prednost nad konkurencijom.

## LITERATURA

Agarwal, A., Biadys, F., & Mckeown, K. R. (2009). Contextual Phrase-Level Polarity Analysis using Lexical Affect Scoring and Syntactic N-grams. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, (pp. 24-32). Stroudsburg, PA, USA.

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman & Hall.

Altrabsheh, N., Cocea, M., & Fallahkha, S. (2014). Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms. *Adaptive and Intelligent Systems - Third International Conference ICAIS 2014*, 8779, 40-49.

Altrabsheh, N., Medhat Gaber, M., & Cocea, M. (2013). SA-E: sentiment analysis for education. *The 5th KES International Conference on Intelligent Decision Technologies (KES-IDT)*, 255.

Analiza emocija, Matematički fakultet u Beogradu. (2011). *Analiza emocija*. Retrieved septembar 10, 2014, from <http://cvetana.mmiljana.com/>

Araque, O., Corcuera, I., Roman, C., Iglesias, C. A., & Sanchez-Rada, F. J. (2015). Aspect based Sentiment Analysis of Spanish Tweets. *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference SEPLN 2015*, (pp. 29-34). Alicante, Spain.

Archak, N., Ghose, A., & Ipeirotis, P. (2007). Show me the money! Deriving the pricing ing power of product features by mining consumer reviews. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*.

Artstein, R., & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34 (4), 555-596.

Aue, A., & Gamon, M. (2005). Customizing Sentiment Classifiers to New Domains: a Case Study. *Proceedings of Recent Advances in Natural Language Processing RANLP*.

Balahur, A. D. (2011). *Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types*. Universidad de Alicante.

Balahur, A., Kozareva, Z., & Montoyo, A. (2009). Determining the Polarity and Source of Opinions Expressed in Political Debates. *Computational Linguistics and Intelligent Text Processing*, 5449, 468-480.

Bansal, M., Cardie, C., & Lee, L. (2008). The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. *Proceedings of COLING*, (pp. 15-18).

Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., & Goldstein, G. (2011). Identifying and Following Expert Investors in Stock Microblogs. *Conference on Empirical Methods in Natural Language Processing*.

Barnes, N. G., Lescault, A. M., & Andonian, J. (2012). Social media surge by the 2012 Fortune 500: Increase use of blogs, Facebook, Twitter and more. *Charlton College of Business Center for Marketing Research*.

Barreda, A. A., Bilgihan, A., Nusair, K., & Okumus, F. (2015). Generating brand awareness in Online Social Networks. *Computers in Human Behavior*, 50, 600-609.

Basuroy, S., Chatterjee, S., & Ravid, A. S. (2003). How critical are critical reviews? the box office effects of film critics, star power and budgets. *Journal of Marketing*, 67 (4).

Bautin, M., Vijayarenu, L., & Skiena, S. (2008). International sentiment analysis for news and blogs. *Proceedings of the International Conference on Weblogs and Social Media ICWSM*.

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. *Marketing science*, 29 (5), 815-827.

Berlanga, F. J., del Jesus, M. J., Gacto, M. J., & Herrera, F. (2006). A Genetic-Programming-Based Approach for the Learning of Compact Fuzzy Rule-Based Classification Systems. *Artificial Intelligence and Soft Computing—ICAISC 2006*, 4029, 182-191.

Berlanga, R., García-Moya, L., Nebot, V., Aramburu, M. J., Sanz, I., & Llidó, D. M. (2015). SLOD-BI: An Open Data Infrastructure for Enabling Social Business Intelligence. *Data Warehousing and Mining*, 11 (4), 1-28.

Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons*, 55 (3), 261-271.

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. *Proceedings of the Association for Computational Linguistics ACL*.

Blue, A. (2015, April 14). *How Twitter Can Help Predict Emergency-Room Visits*. Retrieved November 14, 2015, from UA News - The University of Arizona: <http://uanews.org/story/how-twitter-can-help-predict-emergency-room-visits>

Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2 (1), 1-8.

Borges Tiago, M. T., & Cristovao Verissimo, J. M. (2014). Digital marketing and social media: Why bother? *Business Horizons*, 57, 703-708.

- Bošnjak, Z. (2006). *Inteligentni sistemi i poslovna primena*. Novi Sad: Offset print.
- Bošnjak, Z., Tumbas, P., Bošnjak, S., Grljević, O., & Raković, L. (2010). *Analiza podataka o malim i srednjim preduzećima u Vojvodini inteligentnim metodama i tehnikama*. Subotica: Offset Print Novi Sad.
- Boyd, D. M., & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, 210-230.
- Breck, E., Choi, Y., & Cardie, C. (2007). Identifying expressions of opinion in context. *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI-2007*. Hyderabad, India.
- Broß, J. (2013). *Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques*. Berlin.
- Burde, A. K., Barage, T. G., Shevate, V. B., Suryawanshi, P., & Chapate, V. (2015). Survey on Sentiment Analysis and its Classification Technique. *International Journal of Computer Applications - Proceedings on National Conference on Advances in Computing NCAC*, 5, 21-26.
- Cambria, E., Hussain, A., Durrani, T., & Zhang, J. (2012). Towards a Chinese Common and Common Sense Knowledge Base for Sentiment Analysis. *25th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems*, (pp. 437-446). Dalian, China.
- Chakrabarti, S., Roy, S., & Soundalgekar, M. V. (2003). Fast and accurate text classification via multiple linear discriminant projections. *The VLDB Journal — The International Journal on Very Large Data Bases*, 12 (2), 170-185.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43 (3), 345–354.
- Choi, Y., & Cardie, C. (2008). Learning with compositional semantics as structural inference for subsentential sentiment analysis. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Chow, W. S., & Shi, S. (2015). Investigating Customer's Satisfaction with Brand Pages in Social Networking Sites. *Journal of Computer Information Systems*, 55 (2), 48-58.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1), 37-46.

- Cui, H., Mittal, V., & Datar, M. (2006). Comparative Experiments on Sentiment Classification for Online Product Reviews. In A. Cohn (Ed.), *AAAI'06 proceedings of the 21st national conference on Artificial intelligence - Volume 2* (pp. 1265-1270). AAAI Press ©2006.
- D'Avanzo, E., & Pilato, G. (2015). Mining social network users opinions' to aid buyers' shopping decisions. *Computers in Human Behavior* , 51, 1284–1294.
- Dalal, M. K., & Zaveri, M. A. (2014). Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges. *Applied Computational Intelligence and Soft Computing* , 2014.
- D'Avanzo, E., & Kuflik, T. (2013). E-commerce websites services versus buyers expectations: An empirical analysis of the online marketplace. *International Journal of Information Technology & Decision Making* , 12 (4), 651-677.
- Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. *WWW '03 Proceedings of the 12th international conference on World Wide Web* (pp. 519-528). ACM New York, NY, USA ©2003.
- De Bondt, W. F., & Thaler, R. H. (1994). Financial Decision-Making in Markets and Firms: A Behavioral Perspective. (w4777).
- Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product ratings in revenue forecasting: The case of motion pictures. *Journal of Interactive Marketing* , 21 (4), 23–45.
- Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. *Artificial Intelligence in Medicine* , 64 (1), 17-27.
- Deshpande, M., & Sarkar, A. (2011). BI and Sentiment analysis. *Business Intelligence Journal* , 15 (2), 41-50.
- Devitt, A., & Ahmad, K. (2007). A Lexicon for Polarity: Affective content in financial news text. *Conference on Language for Special Purposes*.
- Dijkmans, C., Kerkhof, P., & Beukeboom, C. J. (2015). A stage to engage: Social media use and corporate reputation. *Tourism Management* , 47, 58-67.
- Ding, X., Liu, B., & Yu, P. S. (2008). A Holistic Lexicon-Based Approach to Opinion Mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08*, (pp. 231-240).
- Đukić, S. (2011). Uloga i značaj društvenih medija u komuniciranju vrednosti marke. *Marketing*, 42, 17-26.

Ellison, N. B., & Boyd, D. M. (2013). Sociality Through Social Network Sites. In W. H. Dutton (Ed.), *The Oxford Handbook of Internet Studies* (pp. 151-173). Oxford University Press.

Feldman, R., & Sanger, J. (2013). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.

Feldman, R., Rosenfeld, B., Bar-Haim, R., & Fresko, M. (2011). The Stock Sonar — Sentiment Analysis of Stocks Based on a Hybrid Approach. *23rd IAAI Conference on Artificial Intelligence*.

Filatova, E. (2012). Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. *The International Conference on Language Resources and Evaluation, LREC*. Istanbul, Turkey.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76 (5), 378-382.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*. Wiley.

Funk, A., Li, Y., Saggion, H., Bontcheva, K., & Leibold, C. (2008). Opinion analysis for business intelligence applications. *OBI '08 Proceedings of the first international workshop on Ontology-supported business intelligence*. New York, NY, USA: ACM.

Gamallo, P., Garcia, M., & Fernández-Lanza, S. (2013). TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets. *Workshop on Sentiment Analysis at SEPLN (TASS2013)*.

Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining Customer Opinions from Free Text. *Advances in Intelligent Data Analysis VI* (Lecture Notes in Computer Science ed., Vol. 3646, pp. 121-132). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ganu, G., Elhadad, N., & Marian, A. (2009). Beyond the stars: Improving rating predictions using Review Text Content. *Twelfth International Workshop on the Web and Databases (WebDB 2009)*, 9, pp. 1-6. Providence, Rhode Island, USA.

Ghiassia, M., Skinnerb, J., & Zimbrea, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, 40 (16), 6266–6282.

Gligorijevic, B., & Luck, E. (2012). Engaging Social Customers – Influencing New Marketing Strategies for Social Media Information Sources. In *Contemporary Research on E-business Technology and Strategy* (pp. 25-40). Springer Berlin Heidelberg.

Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2013). Embodied Affect in Tutorial Dialogue: Student Gesture and Posture. *Proceedings of the 16th International*



*Conference on Artificial Intelligence in Education AIED* (pp. 1-10). Springer-Verlag Berlin Heidelberg.

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Use of Sentiment Analysis for Capturing Patient Experience From Free-Text Comments Posted Online. *Journal of Medical Internet Research* , 15 (11), e239.

Grljević, O., & Bošnjak, Z. Značaj analize sadržaja socijalnih medija. *XVIII YU INFO konferencija*, (pp. 84-89). Kopaonik, Srbija.

Grljević, O., & Šereš, L. (2012). Mogućnosti integracije poslovne inteligencije i socijalnih medija. *XXXIX Simpozijum o operacionim istraživanjima, SYM-OP-IS 2012*.

Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2006). eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business Research* , 59 (4), 449-456.

Hahn, U., Beisswanger, E., Buyko, E., Poprat, M., Tomanek, K., & Wermter, J. (2008). Semantic Annotations for Biology: a Corpus Development Initiative at the Jena University Language & Information Engineering (JULIE) Lab. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008*. Marrakech, Morocco.

Harris, L., & Dennis, C. (2011). Engaging customers on Facebook: Challenges for e-retailers. *Journal of Consumer Behaviour* , 10, 338–346.

Herschel, R. (2011). Business Intelligence: Opportunities for Research. *33rd International Conference on Information Technology Interfaces (ITI)*. Cavtat, Hrvatska.

Hester, J. B., & Vargo, C. J. (2013). Social Network Sites and Social Media: A New Research Paradigm for Strategic Communication? *Proceedings of AAA conference 2013*. Albuquerque, New Mexico.

Hiltbrand, T. (2010). Social Intelligence: The Next Generation of Business Intelligence. *Business Intelligence Journal* , 15 (3), 7-13.

Hopper, A. M., & Uriyo, M. (2015). Using sentiment analysis to review patient satisfaction data located on the internet. *Journal of Health Organization and Management* , 29 (2), 221 - 233.

Hovy, E., & Lavid, J. (2010). Towards a ‘Science’ of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. *International Journal of Translation* , 22 (1), 13-36.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04* (pp. 168-177). New York, NY, USA: ACM.

- Hudson, S., Roth, M. S., Madden, T. J., & Hudson, R. (2015). The effects of social media on emotions, brand relationship quality, and word of mouth: An empirical study of music festival attendees. *Tourism Management* , 47, 68-76.
- Ibrahim, H. S., Abdou, S. M., & Gheith, M. (2015). Sentiment Analysis for Modern Standard Arabic and Colloquial. *International Journal on Natural Language Computing (IJNLC)* , 4 (2), 95-109.
- Jagtap, V., & Pawar, K. (2013). Analysis of different approaches to Sentence-Level Sentiment Classification. *International Journal of Scientific Engineering and Technology* , 2 (3), 164-170.
- Karystianisa, G., Dehghana, A., Kovacevic, A., Keanea, J. A., & Nenadic, G. (2015). Using local lexicalized rules to identify heart disease risk factors in clinical notes. *Journal of Biomedical Informatics* .
- Kasper, W., & Vela, M. (2011). Sentiment Analysis for Hotel Reviews. *Proceedings of the Computational Linguistics-Applications Conference*, (pp. 45–52).
- Kessler, J. S., & Nicolov, N. (2009). Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations. *Proceedings of the Third International AAI Conference on Weblogs and Social Media*, (pp. 90-97).
- Kessler, J. S., Eckert, M., Clark, L., & Nicolov, N. (2010). The 2010 ICWSM JDPa sentiment corpus for the automotive domain. *Proceedings of the 4th International AAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010)*.
- Khadjeh Nassirtoussia, A., Aghabozorgi, S., Ying Waha, T., & Chek Ling Ngob, D. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications* , 41 (16), 7653–7670.
- Khadjeh Nassirtoussia, A., Aghabozorgi, S., Ying Waha, T., & Chek Ling Ngob, D. (2015). Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. *Expert Systems with Applications* , 42 (1), 306–324.
- Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems* , 57, 245–257.
- Kim, J., Choi, D., Hwang, M., & Kim, P. (2014). Analysis on Smartphone Related Twitter Reviews by Using Opinion Mining Techniques. *Advanced Approaches to Intelligent Information and Database Systems, Studies in Computational Intelligence* , 551, 205-212.
- Kim, J., Shaw, E., Wyner, S., Kim, T., & Li, J. (2010). Discerning Affect in Student Discussions. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

- Kim, S. M., & Hovy, E. (2006/1). Automatic identification of pro and con reasons in online reviews. *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions* (pp. 483-490). Morristown, NJ, USA: Association for Computational Linguistics.
- Kim, S.-M., & Hovy, E. (2006/2). Identifying and analyzing judgment opinions. *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference, HLT-NAACL*.
- Koppel, M., & Shtrimberg, I. (2006). Good news or bad news? Let the market decide. *Computing Attitude and Affect in Text: Theory and Applications*, 20.
- Korayem, M., Crandall, D., & Abdul-Mageed, M. (2012). Subjectivity and Sentiment Analysis of Arabic: A Survey. *Advanced Machine Learning Technologies and Applications - the series Communications in Computer and Information Science*, 322, 128-139.
- Korpus savremenog srpskog jezika. (2002). Retrieved novembar 20, 2014, from Matematički fakultet, Univerzitet u Beogradu: <http://korpus.matf.bg.ac.rs/prezentacija/korpus.html>
- Krippendorff, K. H. (2012). *Content analysis: an introduction to its methodology*, 3rd edn. Thousand Oaks: SAGE Publications.
- Ladis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159-174.
- Lerner, E. (2015, Januar 21). *Twitter Can Predict Rates of Coronary Heart Disease, According to Penn Research*. Retrieved Novembar 14, 2015, from Penn News - University of Pennsylvania: <http://www.upenn.edu/pennnews/news/twitter-can-predict-rates-coronary-heart-disease-according-penn-research>
- Litman, D. J., & Forbes-Riley, K. (2004). Annotating Student Emotional States in Spoken Tutoring Dialogues. *5th SIGdial Workshop on Discourse and Dialogue*, (pp. 144-153).
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Liu, B., Hu, M., & Cheng, J. (2005). Opinion Observer: Analyzing and Comparing Opinions on the Web. *WWW '05 Proceedings of the 14th international conference on World Wide Web*, (pp. 342-351).
- Liu, Y., Huang, X., An, A., & Yu, X. (2007). ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs. *SIGIR '07* (pp. 607-614). ACM.
- Maimon, O. Z., & Rokach, L. (2005). *Decomposition Methodology for Knowledge Discovery and Data Mining: Theory and Applications (Machine Perception and Artificial Intelligence)*. World Scientific Pub Co Inc.

- Marketing\_Charts. (2014). *Top Influence on Online Purchase Decisions? Deals and Discounts*. Retrieved januar 2014, from Marketing Charts: <http://www.marketingcharts.com/wp/topics/e-commerce/top-influence-on-online-purchase-decisions-deals-and-discounts-36149/>
- McGlohon, M., Glance, N., & Reiter, Z. (2010). Star Quality: Aggregating Reviews to Rank Products and Merchants. *Fourth International AAAI Conference on Weblogs and Social Media*, (pp. 114-121).
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* , 5, 1093-1113.
- Melnik, M. I., & Alm, J. (2005). Seller reputation, information signals, and prices for heterogeneous coins on eBay. *Southern Economic Journal* , 72 (2), 305–328.
- Mihalcea, R., Banea, C., & Wiebe, J. (2007). Learning multilingual subjective language via cross-lingual projections. *Proceedings of the Association for Computational Linguistics ACL*, (pp. 976-983). Prague, Czech Republic.
- Milošević, N. (2013). <http://inspiratron.org/>. Retrieved septembar 2, 2014, from Sentiment analyzer for Serbian language: <http://inspiratron.org/SerbianSentiment.php>
- Montoyo, A., Martinez-Barco, P., & Blahur, A. (2012). Subjectivity and Sentiment Analysis: An Overview of the current state of the area and envisaged developments. *Decision Support Systems*, 53 (4), 675-679.
- Moraes, R., Valiati, J. F., & Gavião Neto, W. P. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications: An International Journal* , 40 (2), 621-633.
- Mostafā, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications* , 40 (10), 4241–4251.
- Mullich, J. (2012, Septembar 13). *Data Informed - Big Data and Analytics in the Enterprise*. Retrieved Novembar 7, 2015, from Improving the Effectiveness of Customer Sentiment Analysis: <http://data-informed.com/improving-effectiveness-of-customer-sentiment-analysis/#sthash.SX8aYZa0.dpuf>
- Nasukawa, T., & Yi, J. (2003). Sentiment analysis: capturing favorability using natural language processing. *Proceedings of the 2nd international conference on Knowledge capture*, (pp. 70-77).
- Nielsen, R. D., Ward, W., Martin, J. H., & Palmer, M. (2008). Annotating Students' Understanding of Science Concepts. *Proceedings of the Sixth International Conference on Language Resources and Evaluation LREC '08*. Marrakech, Marocco: European Language Resources Association ELRA.

- O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., et al. (2009). Topic-dependent sentiment analysis of financial blogs. *Proceedings of the first international CIKM workshop on topic-sentiment analysis*, (pp. 9-16). Hong Kong, China.
- Osatuyi, B. (2013). Information sharing on social media sites. *Computers in Human Behavior* , 29, 2622-2631.
- Owsley, S., Sood, S., & Hammond, K. J. (2006). Domain specific affective classification of documents. *AAAI Symposium on Computational Approaches to Analysing Weblogs AAAI-CAAW*, (pp. 181-183).
- Pang, B., & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Proceedings of ACL*.
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* , 2 (1-2), 1-135.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, 10, pp. 79-86. Morristown, NJ, USA.
- Paquette, H. (2013). Social Media as a Marketing Tool: A Literature Review. *Major Papers by Master of Science Students* , Paper 2.
- Park, D.-H., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce* , 11 (4), 125-148.
- Parveen, F., Jaafar, N. I., & Ainin, S. (2015). Social media usage and organizational performance: Reflections of Malaysian social media managers. *Telematics and Informatics* , 32, 67-78.
- Pavlou, P. A., & Dimoka, A. (2006). The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information System Research* , 17 (4), 392-414.
- Pawle, J., & Cooper, P. (2006). Measuring emotion e lovemarks, the future beyond brands. *Journal of Advertising Research* , 46 (1), 38-48.
- Phillips, P., Zigan, K., Santos Silva, M. M., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management* , 50, 130-141.
- Pitt, L. F., Berthon, P. R., Watson, R. T., & Zinkhan, G. M. (2002). The Internet and the birth of real consumer power. *Business Horizons* , 45 (4), 7-14.

- Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014). SemEval-2014 Task 4: Aspect Based Sentiment Analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, (pp. 27–35). Dublin, Ireland.
- Pustejovsky, J., & Moszkowicz, J. (2012). The Role of Model Testing in Standards Development: The Case of ISO-Space. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (pp. 3060-3063). Istanbul, Turkey: European Language Resources Association (ELRA).
- Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning - a Guide to Corpus-Building for Applications*. O'Reilly Media Inc.
- Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. *ACL Student Research Workshop*, (pp. 43-48).
- Refaee, E., & Rieser, V. (2014). An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland.
- Reinstein, D. A., & Snyder, C. M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie reviews. *Journal of Industrial Economics* , 53, 27-51.
- Remus, R., Quasthoff, U., & Heyer, G. (2010). SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. *7th International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA).
- Rivero, E. (2014, Februar 26). *Twitter 'big data' can be used to monitor HIV and drug-related behavior, UCLA study shows*. Retrieved Novembar 14, 2015, from UCLA News - University of California, Los Angeles: <http://newsroom.ucla.edu/releases/twitter-big-data-can-be-used-to-250162>
- Sagum, R. A., Vera de, J. G., Lansang, P. J., Narciso, D. S., & Respeto, J. K. (2015). Application of Language Modelling in Sentiment Analysis for Faculty Comment Evaluation. *International MultiConference of Engineers and Computer Scientists 2015 - IMECS 2015, I*.
- Sečujski, M., & Kupusinac, A. (2006). Automatska morfološka anotacija tekstova na srpskom jeziku korišćenjem HMM. *14. Telekomunikacioni forum TELFOR 2006*, (pp. 325-328). Beograd, Srbija.
- Ševa, N., & Kostić, A. (2003). Annotated Corpus and the Empirical Evaluation of Probability Estimates of Grammatical Forms. *Psihologija* , 36 (3), 255-270.

- Shalunts, G., Backfried, G., & Prinz, K. (2014). Sentiment Analysis of German Social Media Data for Natural Disasters. *Proceedings of the 11th International ISCRAM Conference*, (pp. 752-756). Pennsylvania, USA.
- Shen, L. (2015). Exploring Topic and Sentiment of Hotel Reviews. *Advances in Social Science, Education and Humanities Research* , 1021-1024.
- Sianipar, C. P., Putri, R. C., & Wibisono, D. (2012). Social Media Analysis in Performance Measurement: Moving Toward a New Approach. *The 3rd International Conference on Technology and Operations Management - "Sustaining Competitiveness through Green Technology Management"* , (pp. 715-721).
- Siegert, I., Böck, R., & Wendemuth, A. (2014). Inter-rater reliability for emotion annotation in human-computer interaction: comparison and methodological improvements. *Journal on Multimodal User Interfaces* , 8 (1), 17-28.
- Sinclair, J. (2005). Corpus and Text: Basic Principles. In M. Wynne, *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 1-16). Oxford, UK: Oxbow Books.
- Sobkowicz, P., Kaschesky, M., & Bouchard, G. (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly* , 29, 470-479.
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and Trades: the Information Content of Stock Microblogs. *European Financial Management* , 20 (5), 926-957.
- Stanojčić, Ž., & Popović, L. (2002). *Gramatika srpskog jezika*. Beograd: Zavod za udžbenike i nastavna sredstva.
- Steenkamp, M., & Hyde-Clarke, N. (2014). The use of Facebook for political commentary in South Africa. *Telematics and Informatics* , 31 (1), 91-97.
- Stoyanov, V., & Cardie, C. (2008). Annotating topics of opinions. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA).
- Stubbs, A. (2012). Developing Specifications for Light Annotation Tasks in the Biomedical Domain. *Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining*, (p. 71). Istanbul, Turkey.
- Summers, C. (2010). Social Media and Scientific Journals: A Snapshot. *Science Editor* , 33 (3), 75-78.

- Swaminatha, R., Sharma, A., & Yang, H. (2010). Opinion Mining for Biomedical Text Data: Feature Space Design and Feature Selectio. *the Ninth International Workshop on Data Mining in Bioinformatics (BIOKDD 2010)*.
- Takle, P. R., & Gawai, N. (2015). Identification of Student's Behavior in Higher Education from Social Media by using Opinion based Memetic Classifier. *International Journal on Recent and Innovation Trends in Computing and Communication* , 3 (3), 1074 - 1078.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Pearson Education.
- Thackeray, R., Neiger, B. L., Hanson, C. L., & McKenzie, J. F. (2008). Enhancing Promotional Strategies Within Social Marketing Programs: Use of Web 2.0 Social Media. *Health Promot. Pract.* , 9 (4), 338–343.
- Toprak, C., Jakob, N., & Gurevych, I. (2010). Sentence and expression level annotation of opinions in user-generated discourse. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10* (pp. 575-584). Stroudsburg, PA, USA, 2010: Association for Computational Linguistics.
- Tripp, T. M., & Grégoire, Y. (2011). When Unhappy Customers Strike Back on the Internet. *MIT Sloan Management Review* , 52 (3), 37-44.
- Turney, P. D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 417–424). Morristown, NJ, USA: Association for Computational Linguistics.
- Van de Camp, M., & Van de Bosch, A. (2012). The socialist network. *Decision Support Systems*, 53 (4), 761-769.
- Van de Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications* , 42, 4999–5010.
- Varela, P. L., Martins, A. F., Aguiar, P. M., & Figueiredo, M. A. (2013). An Empirical Study of Feature Selection for Sentiment Analysis. *9th Conference on Telecommunications*.
- Vebranka. (2007). *Language resources exploration demo site*. Retrieved Decembar 15, 2014, from Vebranka: <http://hlt.rgf.bg.ac.rs/vebranka/>
- Vincze, V., Szarvas, G., Farkas, R., Móra, G., & Csirik, J. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* , 9 (11), 38-45.



- Waltinger, U. (2010). GermanPolarityClues: A Lexical Resource for German Sentiment Analysis. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*. Valletta, Malta.
- Weiss, S. M., Indurkha, N., & Zhang, T. (2010). *Fundamentals of Predictive Text Mining*. London: Springer-Verlag.
- Wen, M., Yang, D., & Rosé Penstein, C. (2014). Sentiment Analysis in MOOC Discussion Forums: What does it tell us? *The 7th International Conference on Educational Data Mining*, (pp. 130-137).
- Wiebe, J., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., et al. (2002). *NRRC summer workshop on multiple-perspective question answering — final report*. Northeast Regional Research Center.
- Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* , 39 (2-3), 165-210.
- Wiegand, M., & Klakow, D. (2009). The role of knowledge-based features in polarity classification at sentence level. *Proceedings of the 22nd International Florida Artificial Intelligence*. AAAI Press.
- Wiegand, M., Balahur, A., Roth, B., Klakow, D., & Montoyo, A. (2010). A Survey on the Role of Negation in Sentiment Analysis. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, (pp. 60-68).
- Wilson, T. A. (2008). *Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states*. Pittsburgh, USA: University of Pittsburgh and UMI Microform.
- Wilson, T., & Wiebe, J. (2005/1). Annotating attributions and private states. *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky, CorpusAnno '05* (pp. 53-60). Stroudsburg, PA, USA, 2005: Association for Computational Linguistics.
- Wilson, T., & Wiebe, J. (2003). Annotating Opinions in the World Press. *4th SIGdial Workshop on Discourse and Dialogue (SIGdial-03)* , 13-22.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005/2). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05* (pp. 347-354). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Wissler, L., Almashraee, M., Monett, D., & Paschke, A. (2014). The Gold Standard in Corpus Annotation. *5th IEEE Germany Student Conference, IEEE GSC 2014*. Passau, Germany.

- Wu, D., & Butz, C. (2005). On the Complexity of Probabilistic Inference in Singly Connected Bayesian Networks. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 10th International Conference, RSFDGrC 2005, Regina, Canada, August 31 - September 3, 2005, Proceedings, Part I, 3641*, Springer Berlin Heidelberg, 581-590.
- Wyner, S., Shaw, E., Kim, T., Li, J., & Kim, J. (2009). Sentiment Analysis of a Student Q&A Board for Computer Science. *Proceedings of the IJCAI workshop on Computational Models of Natural Argument*.
- Xia, R., Zong, C., Hu, X., & Cambria, E. (2013). Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *IEEE Intelligent Systems* , 28 (3), 10-18.
- Xu, G., Zhang, Y., & Li, L. (2011). *Web mining and social networking - Techniques and applications*. New York, USA: Springer Science+Business Media, LLC.
- Yang, C., Chen, Z., Wang, T., & Sun, P. (2015/1). Research on the Sentiment Analysis of Customer Reviews Based on the Ontology of Phone. *International Conference on Education, Management and Computing Technology*, (pp. 273-280).
- Yang, C.-S., Chen, C.-H., & Chang, P.-C. (2015/2). Harnessing consumer reviews for marketing intelligence: a domain-adapted sentiment classification approach. *Information Systems and e-Business Management* , 13 (3), 403-419.
- Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003). Sentiment Analyzer: Extracting sentiments about a given topic using natural language processing technique. *IEEE International Conference on Data Mining ICDM*, (pp. 427-434).
- Yoo, J., & Kim, J. (2014). Capturing Difficulty Expressions in Student Online Q&A Discussions. *Proceedings of the Twenty-Eighth {AAAI} Conference on Artificial Intelligence* (pp. 208-214). AAAI Press.
- Yunqing, X., Xiaoyu, L., Cambria, E., & Hussain, A. (2014). A Localization Toolkit for Sentic Net. *2014 IEEE International Conference on Data Mining Workshop ICDMW*, (pp. 403 - 408). Šenšen, Kina.
- Zhang, H., & Sheng, S. (2004). Learning Weighted Naive Bayes with Accurate Ranking. *the Fourth IEEE International Conference on Data Mining (ICDM'04)*. Brighton, UK.
- Zhu, F., & Zhang, X. (. (2006). The influence of online consumer reviews on the demand for experience goods: The case of video games. *In International Conference on Information Systems (ICIS)*.

Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie review mining and summarization. *Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM'06* (pp. 43-50). New York, NY, USA: ACM.

Zirn, C., Niepert, M., Stuckenschmidt, H., & Strube, M. (2011). Fine-grained sentiment analysis with structural features. *Proceedings of 5th International Joint Conference on Natural Language Processing*. Chiang Mai, Thailand: Asian Federation of Natural Language Processing.

## PRILOZI

### A. Preuzimanje sadržaja sa sajta *Oceni profesora*

Skript za preuzimanje sadržaja sa sajta *Oceni profesora* kreiran je u Python programskom jeziku primenom sledećeg algoritma:

---

**Algoritam A.1.** Ekstrahovanje sadržaja iz HTML stranica

---

**Input:** Veb stranice sajta *Oceni profesora* (H).

**Output:** Ekstrahovan sadržaj sačuvan u tekstualnim fajlovima (F).

**H** ← skup HTML stranica

**F** ← skup output fajlova

**for each**  $h \in H$  **do**

  otvori URL adresu

  kreiraj  $f \in F$

$p \leftarrow \text{LINKPARSER}(h)$

  upiši  $p$  u  $f$

**return**  $f$

**end for**

**function** LINKPARSER( $h$ )

**for each**  $h \in H$  **do**

**if** tag='span' **and** class='full\_name' **then**

$s += \text{self.recording}$  ← preuzima podatke i dodaje ih prethodno preuzetom sadržaju

**else if** tag='a' **and** class='small\_institution\_name' **then**

$s += \text{self.recording}$

**else if** tag='a' **and** class='small\_institution\_taxonomy\_name' **then**

$s += \text{self.recording}$

**else if** tag='div' **and** class='single\_vote\_block\_datetime' **then**

$s += \text{self.recording}$

**else if** tag='div' **and** class='single\_vote\_block\_comment\_text' **then**

$s += \text{self.recording}$

**else if** tag='a' **and** class='gray\_italic\_dislike\_link' **then**

$s += \text{self.recording}$

**end if**

**return**  $s$

**end for**

**end function**

---

## B. Razdelnik rečenica

Razdelnik rečenica je skript razvijen u Python programskom jeziku primenom algoritma B.1. prikazanog u nastavku. Skript rešava izuzetke sa kojima smo se susreli u radu, a koji se prevashodno odnose na skraćenice, vršeći njihovu zamenu ekvivalentima bez interpunkcijskog znaka na kraju, čime se omogućava nesmetano razdvajanje rečenica. Na ovaj način se uspešno razdvajaju rečenice koje završavaju sa tri ili više tačaka, sa jednim ili više uzvičnika, rečenice koje nemaju interpunkcijski znak na kraju, a rečenice u kojima je prisutna neka od skraćenica ne prelamaju se nakon skraćenice. Skript se sa lakoćom može proširiti prema potrebama kako bi obuhvatio dodatne skraćenice ili izuzetke, a uz manju modifikaciju može se upotrebiti i za razdvajanje teksta po rečima.

---

### Algoritam B.1. Ekstrahovanje sadržaja iz HTML stranica

---

**C** ← skup recenzija  
**F** ← skup output fajlova

```

for all c ∈ C do
    učitaj c
    kreiraj f ∈ F
    p ← SRPSKITOKENIZER(c)
    upiši p u f
return f

```

**end for**

**function** SRPSKITOKENIZER(c)

**V** ← skup izuzetaka u srpskom jeziku, poput npr., dipl. i drugih skraćenica.  
**c\*** ← 0

```

for all v ∈ V do
    if v ∈ c then
        c* = c ∪ {v}
        razdvoj c* prema sledećim pravilima:
        a. Ukoliko tačka sledi iza alfanumeričkog karaktera razdvoj sadržaj.
        b. Razdvoj sadržaj kod pojave skrivenih karaktera: nova linija teksta, tabulator, znak za
           povratak na početak reda i vertikalni tabulator.
        return c*
    end if
end for
end function

```

---

## C. XML shema anotacije

Slika C.1. ilustruje XML shemu na osnovu koje je realizovana anotacija korpusa. Na shemi se vide atributi koji su anotirani: Aspekt, Sentiment polaritet i Intenzitet sa mogućim vrednostima, zatim atributi PozitivnaReč, NegativnaReč, Negacija i Opseg važenja negacije koji nemaju predefinisane vrednosti, već se njihove vrednosti generišu na osnovu sadržaja recenzije.

```

<!ELEMENT Aspekt ( #PCDATA ) >
<!ATTLIST Aspekt OznakaKomentara >
<!ATTLIST Aspekt start #IMPLIED >
<!ATTLIST Aspekt vrednosti ( profesor | predavanja | odnos prema studentima | predmet | materijali |
| | | | | organizacija predmeta | drugi aspekt ) #IMPLIED "profesor" >
<!ATTLIST Aspekt SentimentPolaritet ( pozitivan | negativan | neutralan ) >
<!ATTLIST Aspekt Intenzitet ( 3 | 2 | 1 ) >
<!ATTLIST Aspekt komentar CDATA " " >

<ELEMENT PozitivnaRec (#PCDATA)>
<ELEMENT NegativnaRec (#PCDATA)>

<ELEMENT Negacija (#PCDATA)>
<ELEMENT OpsegVazjenjaNegacije (#PCDATA) >

```

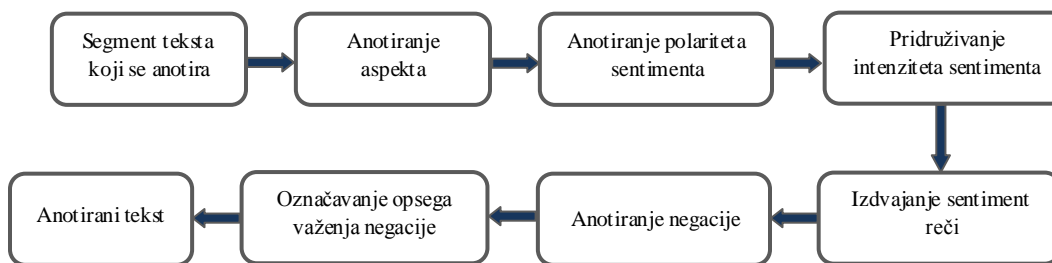
Slika C.1. XML shema anotacije

## D. Uputstvo za anotaciju korpusa

Uputstvo predstavlja vodič za kreiranje sveobuhvatnog i konzistentno anotiranog korpusa za domen visokog obrazovanja. Kroz vodič su date konkretne instrukcije i uputi za identifikovanje, interpretiranje i označavanje različitih aspekata sentimenta definisanih modelom anotacije. Instrukcije i saveti su dati u formi primera i objašnjenja. Svi prezentovani primeri su deo originalnog skupa podataka<sup>48</sup>. Nad skupom podataka izvršena je korekcija pravopisa<sup>49</sup>. Usled kompleksnosti jezika i nepostojanja softvera koji bi automatizovano i bez greške radio korigovanje gramatičkih grešaka, gramatičke greške su zadržane.

Korpus je reprezentovan u formi jednog Microsoft Excel fajla. Korpus je preprocesiran automatskim razdelnikom rečenica tako da jedan red u Excel fajlu predstavlja tačno jednu rečenicu. Rečenicama su pridružene oznake da bi se znalo kojem komentaru pripada rečenica.

Proces anotacije je prikazan na slici D.1.



Slika D.1. Proces anotacije

Pre sprovođenja samog postupka anotacije, anotatori se upućuju na detaljno izučavanje prezentovanog uputstva, u kome je ilustrovana i opisana većina standardnih slučajeva i problematičnih primera. U procesu anotacije anotatori aktivno konsultuju uputstva i koriste ih. Ukoliko se i nakon razmatranja uputstva pojave nejasnoće, anotatori postupaju po sopstvenom nađenju, odnosno odabiraju anotaciju za koju smatraju da je najadekvatnija. Ovakve slučajeve anotatori označavaju u predviđenom polju **Komentar** upisivanjem **procenata pouzdanosti anotacije** na skali od 0% do 99%.

U procesu anotacije pratiti sledeće korake i opšte savete:

1. Pažljivo pročitati rečenicu, ako je potrebno i više puta.
2. Kako se mišljenja ne iskazuju u **formi upita**, ukoliko se pojavi upitna rečenica ne anotirati je.

<sup>48</sup> Primeri prezentovani u uputstvu za anotaciju nisu deo skupa koji anotatori anotiraju.

<sup>49</sup> <https://hacheck.tel.fer.hr/>

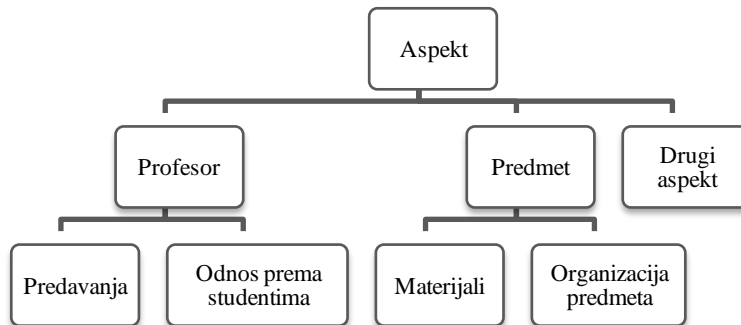
3. Analizirati rečenicu, kako bi se utvrdio predmet o kojem se piše, i označiti rečenicu odabirom jedne od predefinisanih vrednosti za atribut **Aspekt**, slika D.2. Ukoliko rečenica obuhvata više od jednog aspekta, označiti svaki deo rečenice koji se odnosi na jedan aspekt i anotirati svaki deo zasebno (postupak je detaljnije opisan u poglavlju *Postupak anotacije*). Proces anotacije *Aspekta* je prikazan na slici D.3.
4. U obuhvatu rečenice koja je predmet anotacije voditi računa da se **uključ** **interpunkcijski znak** bez praznog polja nakon njega.
5. Anotator utvrđuje da li se o prezentovanom delu teksta, kojem je dodelio adekvatan aspekt, piše u pozitivnom, negativnom ili neutralnom svetlu, na osnovu čega pridružuje polaritet sentimenta postavljanjem vrednosti atributa **SentimentPolaritet**, slika D.4. Ukoliko u nekoj rečenici nema iskazanog sentimenta, takve rečenice treba klasifikovati kao neutralne. Proces anotacije *SentimentPolariteta* je prikazan na slici D.5.
6. Aspekt i polaritet sentimenta mogu biti iskazani na **implicitan i eksplicitan način**. Iz tog razloga anotatori koriste svoja iskustvena saznanja i zdravorazumsko rasuđivanje kako bi utvrdili koji aspekt je predmet pisanja, odnosno koji polaritet sentimenta se pridružuje identifikovanom aspektu.
7. Ukoliko se rečenici ne može pridružiti aspekt ili sentiment bez razmatranja **konteksta**, anotatori se upućuju da pogledaju ostale rečenice koje čine komentar, te da nakon toga pridruže adekvatne atribute rečenici koju anotiraju.
8. Postavljanjem atributa **IntenzitetSentimenta** na vrednost od 1 do 3, anotator pridružuje intenzitet iskazanom sentimentu u tekstu gde 3 označava snažan pozitivan ili negativan sentiment, 2 srednji i 1 slab sentiment.
9. Atributi **PozitivnaReč** i **NegativnaReč**, za razliku od ostalih atributa, nemaju unapred definisanu listu vrednosti sa koje će anotatori odabirati adekvatnu. Naime, da bi se postavila vrednost ovih atributa, reč na osnovu koje je anotator identifikovao sveobuhvatni sentiment iskazan u rečenici, potrebno je pročitati rečenicu i identifikovati jednu ili više takvih reči. Kako su ovakve reči specifične za svaku rečenicu, anotator svaku sentiment reč označava i odabira adekvatnu vrednost **PozitivnaReč** ili **NegativnaReč**, u zavisnosti od sentimenta koju označena reč nosi.
10. Povesti računa o **načinu selektovanja rečenica i reči** kako se ne bi izostavilo slovo ili dodao suvišan prazan prostor (space).
11. Atribut **Negacija** je namenjen označavanju negacija u rečenicama. Svaku reč koja predstavlja negaciju (ne, nema, nije i slično) potrebno je označiti.
12. Atributom **OpsegVaženjaNegacije** obuhvata se deo teksta koji reč negira.

U nastavku je prezentovno uputstvo za svaki atribut definisan modelom anotacije.



## D.1. Atribut Aspekt

Atribut aspekt oslikava predmet komentarisanja. S obzirom na izvor podataka<sup>50</sup>, prirodu sadržaja, kao i način na koji se iznose mišljenja, osnovni aspekti o kojima se piše su *profesor*<sup>51</sup> i *predmet*. Aspekt profesor obuhvata dve kategorije: predavanja i odnos profesora prema studentima, dok aspekt predmet obuhvata: materijale i organizaciju predmeta. Za rečenice čiji se sadržaj ne odnosi na domen visokog obrazovanja, ali koje ipak sadrže određeni sentiment, definisana je vrednost *Drugi aspekt*. Hijerarhija aspekata je predstavljena na slici D.2.



Slika D.2. Vrednosti atributa *Aspekt*

Proces anotiranja *aspekta* je prikazan na slici D.3. Kako stablo odlučivanja prikazuje, anotator prvo pridružuje komentaru neki od aspekata na nižem hijerarhijskom nivou. Ukoliko ne postoji osnova za takav način anotacije, komentaru se dodeljuje aspekt višeg hijerarhijskog nivoa. U nastavku su detaljnije objašnjeni i ilustrovani pojedinačni aspekti.



Slika D.3. Anotiranje atributa *Aspekt* - stablo odlučivanja

<sup>50</sup> Izvor podataka predstavlja sajt <http://oceniprofesora.com/>

<sup>51</sup> U ovom kontekstu profesor se odnosi na sve predavače na visokoškolskim institucijama, odnosno redovne profesore, vanredne profesore, docente, asistente sa doktoratom, asistente, saradnike u nastavi, istraživače pripravnike i drugo osoblje angažovano u nastavnom procesu.

**Predavanja** se odnose na način izlaganja profesora, način prenošenja znanja, koliko dobro profesor objašnjava gradivo koje predaje, stepen interesantnosti predavanja, upotrebljivost predavanja za pripremu ispita, i slično.

Primeri rečenica čiji aspekt je *Predavanje*:

- (P.1) Savršeno predaje i ne postoji mogućnost da nešto ostane nejasno u gradivu.  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.2) Predavanja su retko kad živahna, ali su uvek korisna.  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.3) Predavanja su zanimljiva i nisu bila ništa posebno korisna.  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.4) Njena predavanja nisu nezanimljiva, ali se ne mogu iskoristiti kao podloga za spremanje ispita.  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.5) Na kraju svakog njenog predavanja sam se pitao: Zar je prošlo 2h? ...  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.6) Pošto sam je ukupno pet puta videla na predavanjima, ne znam da li su korisna ili zanimljiva...  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.7) Uvek pripremljena za čas...  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.8) Veoma jednostavno objasni komplikovane stvari.  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.9) Rečenice nedovršene i nemaju puno smisla što onemogućava studenta da vodi beleške na predavanjima.  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]
- (P.10) Ona šta ispredaje, ispredaje i ja retko kada stignem da zapišem nešto u beleške, jer mnogo brzo govori.  
[Aspekt=predavanja; SentimentPolaritet=; Intenzitet=]

Prvih šest rečenica (P.1–P.6) ukazuju na primere u kojima je aspekt *Predavanje* eksplicitno iskazan. Rečenice se razlikuju prema iskazanom sentimentu i njegovom intenzitetu. U narednoj rečenici (P.7) upotrebljen je sinonim za pojam *predavanje*. Rečenice (P.8–P.10) ilustruju situacije u kojima pisac komentara ukazuje na kvalitet i način izlaganja gradiva. Naime, u primerima se ukazuje na stil izražavanja i predavanja profesora.

**Odnos prema studentima** se odnosi na ophođenje profesora prema studentima na nastavi, tokom provere znanja, na konsultacijama i u drugim situacijama, zatim na kvalitet komunikacije između profesora i studenta, (ne)postojanje jasnog kriterijuma pri ocenjivanju, i slično.

Primeri rečenica čiji aspekt je *odnos prema studentima*:

(P.11) U komunikaciji je i više nego korektan.

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

(P.12) Ovakva komunikacija i sposobnosti aktivnog slušanja dovoljno govore o njenim profesionalnim kvalitetima (koji postanu još jasniji ukoliko se analizira na čemu je diplomirala, magistrirala i doktorirala).

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

(P.13) Vraća radove jer u njenom imenu nije bilo dodato prezime Ilić koje sad više nema.

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

(P.14) Koliko ona ulaže u svoj rad, toliko i očekuje od studenta, zato mislim da je opravdano stroga.

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

Prve tri rečenice ilustruju situacije u kojima se na eksplicitan način ukazuje na odnos profesora prema student. Kvalitet komunikacije (P.11) i (P.12) i korektnost odnosa prema studentima (P.13) predstavljaju predmet ovih rečenica. U primeru (P.14) student piše o očekivanjima profesora i ističe strogost kao karakteristiku koja dodatno opisuje njegov odnos prema studentu.

(P.15) Nije toliko sitničav i mnogo je korektniji pri ocenjivanju.

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

(P.16) Na usmenom ako se odgovara među prvima jako korektna.

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

(P.17) Nije otvoren za studentove ideje, tj. ukoliko neki problem ima 2 moguća rešenja, on prihvata samo ono koje je on zamislio, bez obzira da li je drugo rešenje isto tačno.

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

(P.18) Ne ceni trud, samo gleda rezultat.

[Aspekt=*odnos prema studentima*; SentimentPolaritet=; Intenzitet=]

Navedeni primeri ilustruju odnos profesora prema studentima u procesu provere znanja. U prva dva primera (P.15) i (P.16) predmet pisanja je način ispitivanja i ocenjivanja studenata. U primeru (P.17) student ukazuje na želju za širim shvatanjima profesora i otvorenijim odnosom. U pozadini primera (P.18) je iskazana sugestija za uvažavanjem angažmana studenta na predmetu prilikom ocenjivanja.

(P.19) Jednom je bitno odakle je Egej, a sledeći put nije bitno kako se zove Antigonin brat.

[Aspekt= odnos prema studentima; SentimentPolaritet=; Intenzitet=]

(P.20) Razumem da radite stresan posao, ali mora se imati neka samokontrola.

[Aspekt= odnos prema studentima; SentimentPolaritet=; Intenzitet=]

(P.21) Jedini problem je u tome što ta energija nekad može da ima rezidual u vidu besa i malo agresivnijeg ponašanja :).

[Aspekt= odnos prema studentima; SentimentPolaritet=; Intenzitet=]

(P.22) Njena filozofija je ja sam asistent i tu sam da vam svima zagorčam život...

[Aspekt= odnos prema studentima; SentimentPolaritet=; Intenzitet=]

(P.23) Kod nje samo obavezni predmeti, od izbornih bežite.

[Aspekt= odnos prema studentima; SentimentPolaritet=; Intenzitet=]

(P.24) Ne sme niko ništa da je pita,..... inače

[Aspekt= odnos prema studentima; SentimentPolaritet=; Intenzitet=]

U primeru (P.19) student ukazuje na nejasno postavljene kriterijume na predmetu, koji mogu dovesti do zabune, nezadovoljstva i nesigurnosti studenata. Naredni primeri (P.20–P.23) ilustruju situacije u kojima se implicitno ukazuje na odnos profesora prema studentima. U primeru (P.24) u pozadini iskazanog mišljenja je strah studenata koji posledično ograničava njihovu komunikaciju sa profesorom.

Ukoliko se rečenica koja je predmet anotacije odnosi na profesora, ali se ne može kategorisati u jednu od prethodne dve kategorije (predavanja i odnos prema studentima), takvoj rečenici dodeliti aspekt **Profesor**. Primeri rečenica čiji aspekt je *profesor*:

(P.25) Najčešće ne zna šta priča i predaje.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

(P.26) Trudi se da da svoj maksimum na svakom času.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

(P.27) Ne vlada materijom, jako nesigurna.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

(P.28) Jeste malo kapriciozan, ali je ipak dobar profesor.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

U navedenim primerima (P.25–P.28) studenti opisuju različite karakteristike profesora.

(P.29) Sve u svemu, pozitivna ocena.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

(P.30) Suviše nonšalantan.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

(P.31) Ponašanje neprimereno.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

Kako je priroda sajta sa kojeg su podaci prikupljeni takva da studenti prvenstveno iskazuju svoje mišljenje o konkretnom profesoru, ukoliko student iskaže svoj stav ili ocenu, ali ne naznači eksplicitno aspekt, poput primera (P.29), može se pretpostaviti da je reč o sveukupnoj oceni profesora. Primeri (P.30) i (P.31) ukazuju na situacije u kojima student nije ni na implicitan ni na eksplicitan način ukazao na precizniji aspekt pisanja, naime, ne zna se da li je profesor nonšalantan u odnosima sa studentima ili tokom svojih predavanja, odnosno ne zna se kada i u kojim situacijama se profesor neprimereno vlada, ali je jasno da se upućene kritike, odnosno pohvale odnose na profesora, te ovakvim primerima treba dodeliti aspekt *profesor*.

(P.32) Šteta da je jedan ovakav profesor otišao u penziju, bila mi je čast ...

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

(P.33) Samo reči hvale za ovog predavača, sve je bilo vrhunski.

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

(P.34) Bez konkurencije jedan od najlošijih asistenata na faksu..

[Aspekt=profesor; SentimentPolaritet=; Intenzitet=]

U navedenim primerima (P.32–P.34) studenti eksplicitno govore o profesorima, ali bez preciznijeg navođenja aspekta (odnos prema studentima ili predavanja).

**Materijali** se odnose na dostupnost i kvalitet nastavnih materijala. Primeri rečenica čiji aspekt je *materijali*:

(P.35) Najteži ispit zbog nepodesnog materijala za učenje.

[Aspekt=materijali; SentimentPolaritet=; Intenzitet=]

(P.36) Materijal za učenje jeste dostupan, ali je nerazumljiv, nabacan, bez reda, bez nekog razumljivog rečnika.

[Aspekt=materijali; SentimentPolaritet=; Intenzitet=]

(P.37) Materijali nisu baš zavidnog kvaliteta.

[Aspekt=materijali; SentimentPolaritet=; Intenzitet=]

(P.38) Takođe, zadaci su vrlo zastareli i nekorisni, mahom svi liče jedan na drugi.

[Aspekt=materijali; SentimentPolaritet=; Intenzitet=]

U primerima (P.35–P.37) studenti eksplicitno ističu aspekt predmet u svojim komentarima, dok u primeru (P.38) student piše o kvalitetu zadataka na osnovu kojih se priprema ispit, te na implicitan način ukazuje na kvalitet dostupnih materijala.

**Organizacija predmeta** se odnosi na način na koji je nastava organizovana, logistiku ispita, kako je npr. predmet podeljen na kolokvijume i na slične tehničke aspekte izvođenja nastave i provere znanja. Primeri rečenica čiji aspekt je *organizacija predmeta*:

(P.39) Smatram da je sama organizacija nastave i polaganja ispita izuzetno loša.

[Aspekt=organizacija predmeta; SentimentPolaritet=; Intenzitet=]

(P.40) Ne znam kakva je u NS, verovatno je ceo predmet tamo organizovan na bolji način ali u SU nije baš bajno pa zato malo slabije ocene.

[Aspekt=organizacija predmeta; SentimentPolaritet=; Intenzitet=]

(P.41) Sistem rada na ovom predmetu je odličan.

[Aspekt=organizacija predmeta; SentimentPolaritet=; Intenzitet=]

(P.42) Svaki ispit je podeljen u glave, pri čemu se svaka glava polaže kao zaseban ispit, pri čemu ispitivanje traje i po više od 3, 4 sata...

[Aspekt=organizacija predmeta; SentimentPolaritet=; Intenzitet=]

(P.43) Ne da Roska ocenu bez kupljene knjige. :D.

[Aspekt=organizacija predmeta; SentimentPolaritet=; Intenzitet=]

U primerima (P.39) i (P.40) studenti u svojim komentarima eksplicitno spominju organizaciju predmeta. U primeru (P.41) sistem rada ukazuje na to da je reč o kompletnoj organizaciji predmeta, dok se u primeru (P.42) govori o generalnoj organizaciji svih ispita kod određenog profesora. Primer (P.43) ilustruje implicitno iskazivanje ovog aspekta.

Ukoliko se u rečenici koja je predmet anotacije piše o predmetu koji profesor ili asistent predaje, ali ne i o nastavnim materijalima, odnosno organizaciji predmeta, takvoj rečenici je potrebno dodeliti aspekt **Predmet**. Primeri rečenica čiji aspekt je *predmet*:

(P.44) Tolika su mi bila očekivanja od ovog predmeta, a dobila sam samo nepotreban stres.

[Aspekt=predmet; SentimentPolaritet=; Intenzitet=]

(P.45) Ceo predmet je bizaran, na kraju nemate nijedan pristojan rad, a nemate ni novo znanje.

[Aspekt=predmet; SentimentPolaritet=; Intenzitet=]

(P.46) Jedino što se da naučiti iz ovog predmeta je da ne treba kasniti i piti vodu pored kompjutera.

[Aspekt=predmet; SentimentPolaritet=; Intenzitet=]

(P.47) Totalno beskoristan predmet prirodne nauke 3na ovom fakultetu.

[Aspekt=predmet; SentimentPolaritet=; Intenzitet=]

Ukoliko u rečenici koja je predmet anotacije dođe do **promene aspekta**, anotiranje raditi parcijalno, odnosno anotirati deo rečenice koji se odnosi na jedan aspekt i postupak ponoviti sa svakom promenom aspekta u rečenici. Primeri rečenica sa više od jednog aspekta:

- (P.51) Samo čita sa papira, i traži da se sve uči napamet.  
[Aspekt=predavanja; odnos prema studentima; SentimentPolaritet=; Intenzitet=]
- (P.52) Izuzetno dobar predavač, svakako autoritativan, ali i korektan za komunikaciju kada mu se priđe sa jasnim pitanjem ili zahtevom.  
[Aspekt= profesor; odnos prema studentima; SentimentPolaritet=; Intenzitet=]
- (P.53) Predmet je vrlo koristan, profesorka je odlična, poseduje veliko znanje i ume da ga prenese na studente.  
[Aspekt= predmet; profesor; predavanja; SentimentPolaritet=; Intenzitet=]

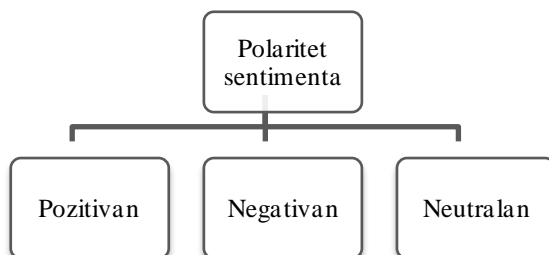
Tokom anotacije se može pojaviti bilo koja kombinacija navedenih aspekata, potrebno je da anotatori anotiraju deo rečenice do tačke u kojoj nastaje promena aspekta pisanja, zatim da od promene aspekta ponove postupak anotacije. U primeru (P.51) student u prvom delu rečenice ukazuje na kvalitet izlaganja profesora, dok u drugom delu govori o odnosu profesora prema studentu tokom provere znanja. Student u primeru (P.52) u prvom delu ističe svoje mišljenje o profesoru, nakon čega opisuje odnos profesora prema studentima. U poslednjem primeru (P.53) student prvo govori o kvalitetu predmeta, potom o kvalitetu profesora i njegovoj sposobnosti da prenese znanje.

Ukoliko je emotikon deo rečenice (nalazi se na početku, sredini ili kraju rečenice), ne odvajati ga iz anotacije. U rečenici „Super profesor :)“ potrebno je celoj rečenici dodeliti aspekt profesor. Ne odvajati ovakve primere na dve anotacije, gde će se posebno anotirati tekst „super profesor“, a posebno „:)“. Emotikone zasebno anotirati samo kada oni samostalno predstavljaju rečenicu.

Sadržaj rečenica koje su predmet anotacije može izaći iz okvira visokog obrazovanja, na primer: lepota profesora, saveti studenata u kojima se ne opisuje nijedan od gorenavedenih aspekata, rečenice u kojima je predmet pisanja bez uticaja na (ne)kvalitet obrazovnog procesa, predavača i generalno visoko obrazovanje. Takvim rečenicama se dodeljuje vrednost **Drugi aspekt**. Ovakve rečenice, mada se ne odnose na domen visokog obrazovanja, mogu nositi sentiment, te im je potrebno pridružiti i ostale atribute (sentiment polaritet, intenzitet, pozitivne i negativne reči i negaciju ukoliko je ima u rečenici).

## D.2. Atribut Sentiment polaritet

U ovom koraku anotacije, u rečenicama u kojima su iskazana uverenja, emocije ili mišljenja, u kojima pisac procenjuje ili sudi o predmetu pisanja, potrebno je prepoznati sentiment (emociju) koji pisac iskazuje u svom komentaru. Atribut *Sentiment polaritet* može poprimiti jednu od sledećih vrednosti: *pozitivan*, *negativan*, *neutralan*, slika D.4.



Slika D.4. Vrednosti atributa *Sentiment polaritet*

U definisanju polariteta sentimenta, anotatori mogu razmotriti sledeći kriterijum:

Da li je pisac u delu komentara, kojem je u prethodnom koraku pridružen aspekt, iskazao pozitivnu emociju, te se predmet o kojem piše može smatrati kao prednost ili poželjna osobina, ili je iskazao negativnu emociju, te se predmet o kojem se piše može smatrati nedostatkom ili nepoželjnom osobinom?

Ukoliko je prvo ili drugo tačno, može se anotirati pozitivan ili negativan sentiment, a ukoliko je netačno, vrednost atributa *Sentiment polaritet* se može postaviti na *neutralan*. Proces anotiranja je prikazan na slici D.5.



Slika D.5. Anotiranje atributa *Sentiment polaritet* – stablo odlučivanja



U nastavku su dati primeri koji detaljnije ilustruju različite vrednosti ovog atributa.

(P.54) Jedan od najboljih profesora.

[Aspekt=; SentimentPolaritet=pozitivan; Intenzitet=]

(P.55) Profesor koji se trudi da mu predavanja budu zanimljiva.

[Aspekt=; SentimentPolaritet=pozitivan; Intenzitet=]

(P.56) Takođe, profesor sa izgrađenim kriterijumom, nema veze ko izlazi koji put, koji je rok, nema uticaja raspoloženje, pravi profesionalac i to možda nekima smeta jer nema tu na smeh, lepe oči i sl. već se traži znanje i mnogo bitnije, razmišljanje.

[Aspekt=; SentimentPolaritet=pozitivan; Intenzitet=]

(P.57) U svakom slučaju, čast mi je bila da slušam i polažem kriminologiju, kao veoma stručan, značajan i zanimljiv predmet, kod profesora Kesića.

[Aspekt=; SentimentPolaritet=pozitivan; Intenzitet=]

Primeri (P.54–P.57) ilustruju rečenice u kojima su iskazana **pozitivna** osećanja. Na osnovu primera može se zaključiti da su karakteristike profesora, pristup studentima, kvalitet predmeta i drugi aspekti o kojima se piše poželjne osobine koje studenti cene.

(P.58) Mnogo je dosadna.

[Aspekt=; SentimentPolaritet=negativan; Intenzitet=]

(P.59) Žena jako loše predaje, ne zna da drži pažnju uopšte.

[Aspekt=; SentimentPolaritet=negativan; Intenzitet=]

(P.60) Sklona je tome da uzima studente na zub.

[Aspekt=; SentimentPolaritet=negativan; Intenzitet=]

(P.61) Što se tiče materijala za učenje stvarno bi bilo poželjno da odrade neku zbirku zadataka, sa postupcima ili barem rešenjima zadataka, da bi studenti znali šta se očekuje od njih.

[Aspekt=; SentimentPolaritet=negativan; Intenzitet=]

(P.62) Studenti su dolazili u fazu da uče takve detalje ne shvatajući više šta je suština dela.

[Aspekt=; SentimentPolaritet=negativan; Intenzitet=]

(P.63) Predugo trajanje usmenog ispita (9-17).

[Aspekt=; SentimentPolaritet=negativan; Intenzitet=]

Primeri (P.58–P.63) ilustruju situacije u kojima pisac komentara ističe **negativnu** stranu i nepoželjne karakteristike aspekta o kojem piše.

(P.64) Slažem se za skripte sa kolegama dole.

[Aspekt=; SentimentPolaritet=neutralan; Intenzitet=]

(P.65) Predavanja su obavezna...

[Aspekt=; SentimentPolaritet=neutralan; Intenzitet=]

(P.66) Po njenom sistemu se može skupiti max 120 bodova - 95 je potrebno za 10.

[Aspekt=; SentimentPolaritet=neutralan; Intenzitet=]

(P.67) Stalno ističe da je došao sa FTN.

[Aspekt=; SentimentPolaritet=neutralan; Intenzitet=]

Primeri (P.64) i (P.65) ilustruju primere u kojima pisac komentara ne iskazuje jasno emociju, te se ovakve rečenice anotiraju kao **neutralne**. Ukoliko u rečenici student iskazuje **činjenicu**, ne svoje subjektivno mišljenje ili stav po pitanju profesora, odnosno predmeta koji profesor predaje (P.66) i (P.67), takvim rečenicama se takođe dodeljuje neutralan sentiment.

Rečenicama u kojima pisac teksta govori o jednom aspektu, ali iskazuje dva suprotna sentimenta, potrebno je tako anotirati da se svakom delu rečenice dodeli adekvatan polaritet sentimenta.

(P.68) **Materijal za učenje jeste dostupan**, ali je nerazumljiv, nabacan, bez reda, bez nekog razumljivog rečnika.

[Aspekt=materijal; materijal; SentimentPolaritet=pozitivan; negativan; Intenzitet=]

U primeru (P.68) student govori o materijalu, međutim, u prvom delu rečenice iskazuje pozitivan sentiment o dostupnosti materijala, dok u drugom delu rečenice govori o njihovom nezadovoljavajućem kvalitetu. Ovakve rečenice je potrebno anotirati u dva koraka, tako da se prvom delu dodeli pozitivan sentiment, dok se drugom delu dodeljuje negativan (postupak anotiranja ilustruju slike D.15. i D.16).

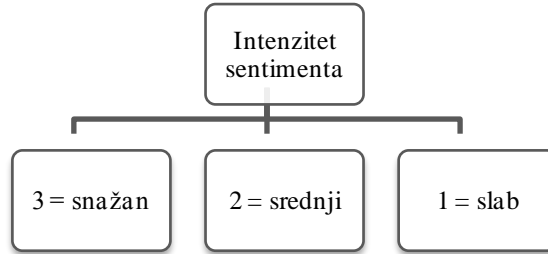
Ukoliko rečenica sadrži dva ili više aspekata, potrebno je svakom aspektu pridružiti adekvatan sentiment (postupak anotiranja ilustruju slike D.15 i D.16, (P.69).

(P.69) **Iako ovaj profesor odlično zna svoj posao**, loš je pedagog i **apsolutno nema komunikaciju sa studentima**.

[Aspekt=profesor; profesor; odnos prema studentima; SentimentPolaritet=pozitivan; negativan; negativan; Intenzitet=]

### D.3. Atribut Intenzitet sentimenta

Iskazani sentiment u rečenici može biti pozitivan, negativan i neutralan. Kako se moglo primetiti iz prethodnih primera, jačina pozitivnog ili negativnog osećanja nije podjednaka u svim komentarima, te je potrebno anotirati i njegov intenzitet. Intenzitet iskazanog osećanja može biti snažan, srednji i slab, slika D.6. U anotaciji će se koristiti numeričke oznake 3 za snažan, 2 za srednji i 1 za slab.



Slika D.6. Vrednosti atributa *Intenzitet sentimenta*

Ukoliko u svojim komentarima studenti govore u superlative, vrednost atributa *IntenzitetSentimenta* može se postaviti na 3, primeri (P.70–P.77). Ukoliko ističu pozitivnu ili negativnu osobinu bez posebnog oduševljenja ili euforije, vrednost ovog atributa može se postaviti na 2, primeri (P.78–P.84), dok se vrednost 1 postavlja za slab intenzitet osećanja, primeri (P.85–P.89). Neutralni sentiment ne može da ima intenzitet jer nema ni pozitivnog ni negativnog osećanja. Ukoliko osećate potrebu da dodelite intenzitet rečenici koju ste okarakterisali kao neutralnu, zapitajte se da li je zaista neutralna.

(P.70) Jedan od najboljih profesora.

[Aspekt=; SentimentPolaritet=; Intenzitet=3]

(P.71) Takođe, profesor sa izgrađenim kriterijumom, nema veze ko izlazi koji put, koji je rok, nema uticaja raspoloženje, pravi profesionalac i to možda nekima smeta jer nema tu na smeh, lepe oči i sl. već se traži znanje i mnogo bitnije, razmišljanje.

[Aspekt=; SentimentPolaritet=; Intenzitet=3]

(P.72) U svakom slučaju, čast mi je bila da slušam i polažem kriminologiju, kao veoma stručan, značajan i zanimljiv predmet, kod profesora Kesića.

[Aspekt=; SentimentPolaritet=; Intenzitet=3]

(P.73) Žena jako loše predaje, ne zna da drži pažnju uopšte.

[Aspekt=; SentimentPolaritet=; Intenzitet=3]

(P.74) U komunikaciji je i više nego korektan.

[Aspekt=; SentimentPolaritet=; Intenzitet=3]

(P.75) Izuzetno obrazovan i inteligentan.

- [Aspekt=; SentimentPolaritet=; Intenzitet=3]
- (P.76) Gradivo i materijali za učenje su preobimni.  
[Aspekt=; SentimentPolaritet=; Intenzitet=3]
- (P.77) Vrlo je neprijatna na ispitu, svojim komentarima ponižava studente i može se desiti da dobijete jedno književno delo, a da vam postavi zbog sličnosti sa tim delom neko sasvim drugo i očekuje da odgovarate to drugo delo koje nije vaše ispitno pitanje.  
[Aspekt=; SentimentPolaritet=; Intenzitet=3]
- (P.78) Profesor koji se trudi da mu predavanja budu zanimljiva.  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.79) Mnogo je dosadna.  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.80) Studenti su dolazili u fazu da uče takve detalje ne shvatajući više šta je suština dela.  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.81) Predugo trajanje usmenog ispita (od 9-17).  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.82) Studenti se dovode u zabludu.  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.83) Ne želi definicije da se znaju, već mu je važnije da se gradivo razume.  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.84) Bilo je sve jasno, zanimljivo, i cool. :)  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.85) Sklona je tome da uzima studente na zub.  
[Aspekt=; SentimentPolaritet=; Intenzitet=2]
- (P.86) Što se tiče materijala za učenje stvarno bi bilo poželjno da odrade neku zbirku zadataka, sa postupcima ili barem rešenjima zadataka, da bi studenti znali šta se očekuje od njih.  
[Aspekt=; SentimentPolaritet=; Intenzitet=1]
- (P.87) Jedina mala zamerka je da ne luta.  
[Aspekt=; SentimentPolaritet=; Intenzitet=1]
- (P.88) Predavanja možda nisu interesantna, ali takvo je gradivo.  
[Aspekt=; SentimentPolaritet=; Intenzitet=1]
- (P.89) Retko ko dobije objašnjenje zašto nešto valja ili ne valja i zašto je dobio određenu ocenu.  
[Aspekt=; SentimentPolaritet=; Intenzitet=1]

#### D.4. Atribut Sentiment reč

Kada je identifikovan pozitivan ili negativan sentiment u rečenici koja je predmet anotacije, anotatori se upućuju da prepoznaju reč, odnosno reči u rečenici koje nose identifikovano osećanje. Sentiment reč je potrebno označiti i u zavisnosti od njenog polariteta odabrati adekvatnu vrednost: *PozitivnaReč* ili *NegativnaReč*. U izdvajanju reči koja nosi sentiment voditi se **MIN strategijom**, koja podrazumeva označavanje minimalne jedinice teksta koja nosi sentiment, sem u izuzetnim slučajevima kada se proceni da jedna reč nedovoljno opisuje pozadinski sentiment – tada se može označiti više od jedne pozitivne ili negativne reči (primeri: zna, ne zna, svaka čast, i sl.). Radi dodatnog pojašnjenja samog postupka anotacije, pogledati [Prilog D.7](#).

Primeri rečenica sa izdvojenim sentiment rečima:

(P.90) Jedan od najboljih profesora.

[Aspekt=; SentimentPolaritet=; Intenzitet=; PozitivnaRec=najboljih]

(P.91) U svakom slučaju, čast mi je bila da slušam i polažem kriminologiju, kao veoma stručan, značajan i zanimljiv predmet, kod profesora Kesića.

[Aspekt=; SentimentPolaritet=; Intenzitet=; PozitivnaRec=strucan, znacajan, zanimljiv]

(P.92) U komunikaciji je i više nego korektan.

[Aspekt=; SentimentPolaritet=; Intenzitet=; PozitivnaRec=korektan]

(P.93) Žena jako loše predaje, ne zna da drži pažnju uopšte.

[Aspekt=; SentimentPolaritet=; Intenzitet=; NegativnaRec=loše, ne zna]

(P.94) Gradivo i materijali za učenje su preobimni.

[Aspekt=; SentimentPolaritet=; Intenzitet=; NegativnaRec = preobimni]

(P.95) Mnogo je dosadna.

[Aspekt=; SentimentPolaritet=; Intenzitet=; NegativnaRec=dosadna]

(P.96) Studenti se dovode u zabludu.

[Aspekt=; SentimentPolaritet=; Intenzitet=; NegativnaRec=zabludu]

(P.97) Jedina mala zamerka je da ne luta.

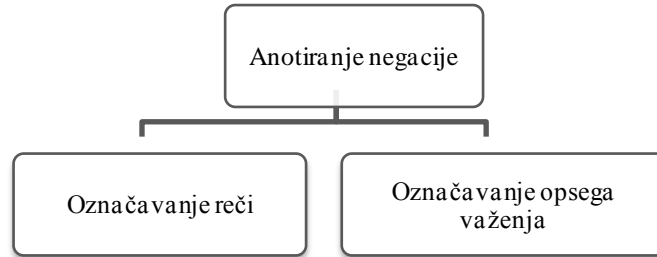
[Aspekt=; SentimentPolaritet=; Intenzitet=; NegativnaRec=zamerka, luta]

Opšte napomene o označavanju sentiment reči:

1. U jednoj rečenici može biti **više od jedne** sentiment reči. Označiti svaku ponaosob.
2. Kao sentiment reč ne izdvajati **lična imena**, primer *Boban Stakić* – nije validna sentiment reč.
3. Ne izdvajati **povratne zamenice** kao deo sentiment reči, primer *razume se* nije validna sentiment reč, umesto toga označiti samo *razume*.
4. **Emotikoni/smajliji** takođe se mogu anotirati kao pozitivna ili negativna sentiment reč.

## D.5. Anotiranje negacije

U detektovanju i označavanju negacije potrebno je koncentrisati se samo na negacije glagola, odnosno negacije koje su odvojene od ostatka teksta. Kako je prikazano na slici D.7, potrebno je anotirati: *reč* koja označava negaciju i *opseg važenja* negacije. Opseg važenja negacije se odnosi na set reči čije značenje negacija menja.



Slika D.7. Postupak anotacije negacije

U označavanju negacije i njenog opsega važenja voditi se **MIN/MAX** strategijom:

**MIN strategija** se odnosi na označavanje negacije, kao što je već istaknuto potrebno je izdvojiti reč (primeri *ne*, *nema*, *nije*).

**MAX strategija** se primenjuje u označavanju opsega važenja negacije. Potrebno je označiti maksimalan set reči na koje negacija koja prethodi ima uticaj, odnosno čije značenje menja. Ne označavati kompletan tekst do kraja rečenice, nego je potrebno označavati samo deo teksta koji je pod uticajem negacije!

(P.98) Profesorka koja **ne zna da predaje**, mrmrlja u bradu i **ne ume da drži disciplinu** i stoga oni koji nešto žele i da čuju, **ne mogu**.

Primer (P.98) pre svega ilustruje prisustvo višestruke negacije u jednoj rečenici, konkretno, tri puta se pojavljuje rečca *ne*. Svako pojavljivanje negacije u rečenici potrebno je odvojeno anotirati sa pripadajućim opsezima važenja – u primeru (P.98) opseg važenja negacije je podvučen. Reč, odnosno rečca, koja je već jednom označena kao negacija ne bi smela da se pojavi u opsegu važenja druge negacije. Prva negacija koja se javlja u rečenici – *ne*, menja značenje teksta *zna da predaje*, što bi ujedno predstavljalo maksimalni deo teksta koji je pod uticajem ove negacije. Dakle, ne uključivati u opseg negacije i suvišne delove rečenice koji nisu pod uticajem negacije, u konkretnom primeru to bi bio deo teksta *mrmrlja u bradu*.

Opseg važenja negacije obično se prostire do interpunkcijskog znaka, nekog veznika ili celine u rečenici. Kao što se može primetiti u prethodnom primeru, kod zareza prestaje važenje prve negacije, drugoj negaciji opseg važenja prestaje kod veznika *i*, a trećoj sa tačkom na kraju rečenice.

## D.6. Alat za anotaciju

Za potrebe anotacije koristiće se MAE (Multi-document Ananotation Environment) alat<sup>52</sup>. U nastavku su data detaljna uputstva za instaliranje i pokretanje MAE alata.

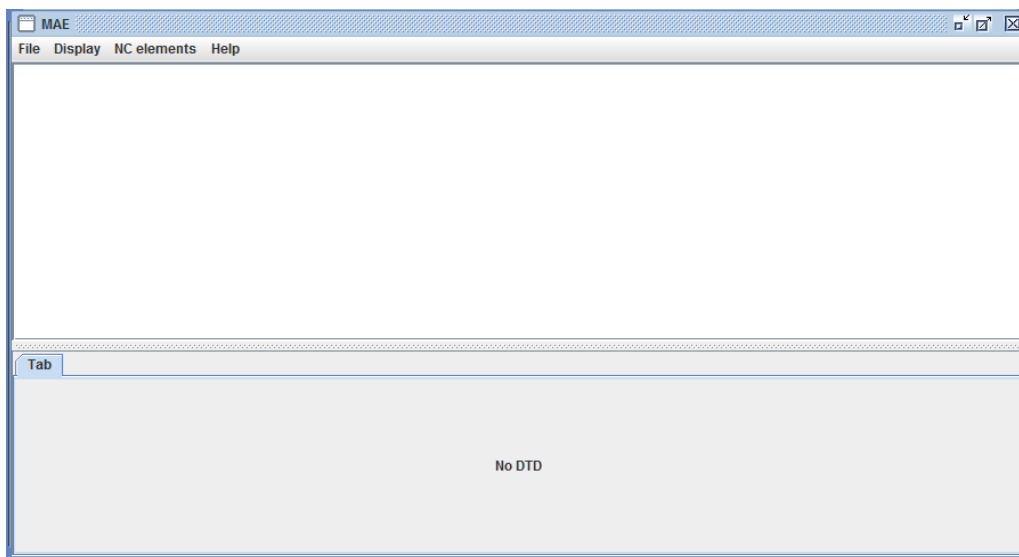
1. Za funkcionisanje alata potrebno je instalirati Javu. Instalacija Jave se može preuzeti sa sledećeg linka <https://java.com/en/download/>. Izgled stranice za preuzimanje je prikazan na slici D.8. Potrebno je odabrati (kliknuti) na **Free Java Download**, na narednom prozoru koji se pojavi odabrati **Agree and Start Free Download**, kada će se pokrenuti preuzimanje Jave. Preuzeta instalacija će imati **.exe** ekstenziju, potrebno je dva puta kliknuti na ovaj fajl, čime se pokreće instalacija Jave na Vašem računaru.



Slika D.8. Preuzimanje Jave

2. Svaki anotator će dobiti folder `MAE_v0.9.6`, koji je potrebno iskopirati i sačuvati na D particiji računara.
3. MAE alat se pokreće duplim klikom na `mae_v0.9.6.jar` fajl. Slika D.9. daje prikaz radnog okruženja alata MAE.

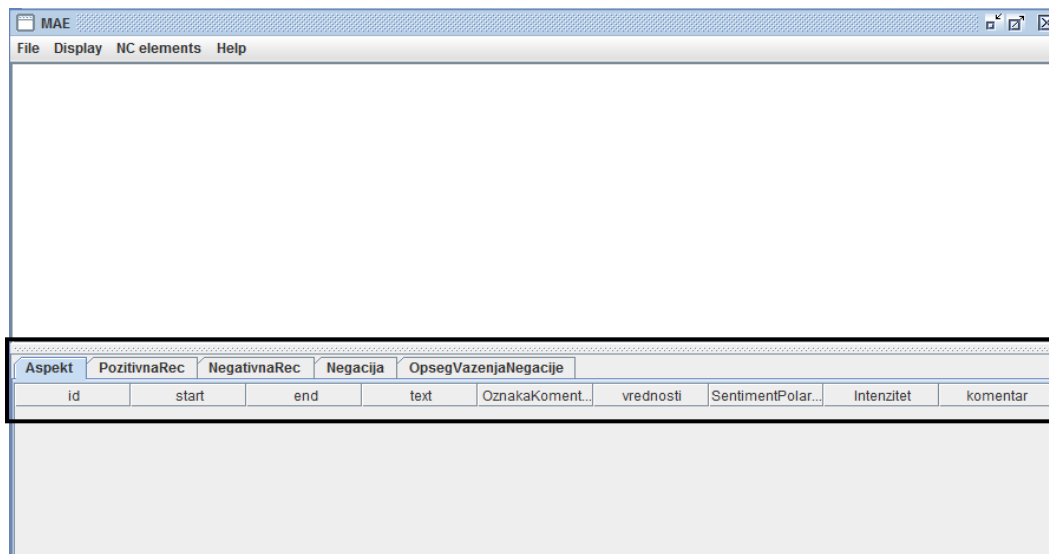
<sup>52</sup> Alat je dostupan na: <http://amberstubbs.net/mae.php>



Slika D.9. Izgled radnog okruženja MAE alata

- U folderu MAE\_v0.9.6 svaki anotator će pronaći folder sa svojim imenom. Unutar foldera se nalaze sledeći fajlovi:
  - XML fajl pod nazivom *dtdSchema* i
  - tekstualni** fajl sa komentarima.
- Pri svakom pokretanju MAE radnog okruženja potrebno je odabrati sa menija **File** opciju **LOAD DTD**, kada će se otvoriti MAE folder, gde je potrebno iz foldera sa Vašim imenom odabrati *dtdShema* fajl i odabrati **OPEN** dugme.

Slika D.10. prikazuje MAE radno okruženje nakon učitavanja fajla dtdSchema, gde se može uočiti da su učitani svi prethodno definisani i opisani atributi anotacije.

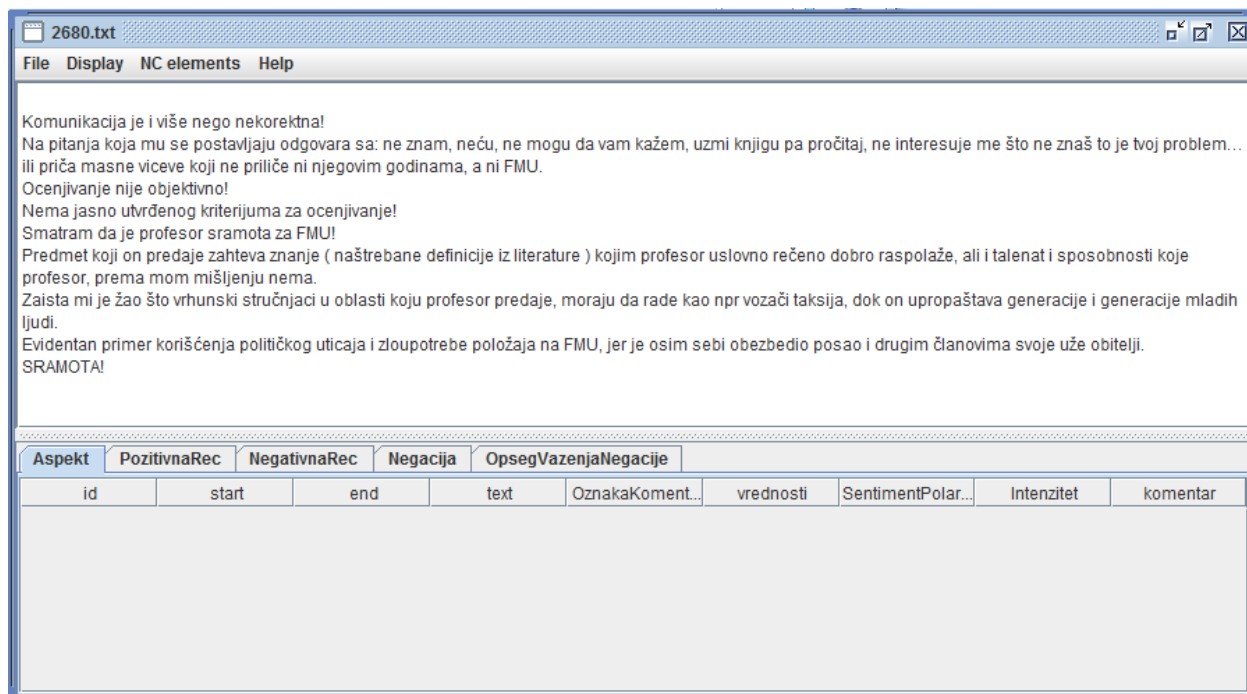


Slika D.10. Radno okruženje nakon učitavanja DTD sheme



- U narednom koraku je potrebno učitati jedan po jedan *tekstualni* fajl nad kojim će se raditi anotacija. Sa menija **File** odabrati **LOAD FILE**, iz foldera odabrati tekstualni fajl i odabrati dugme **OPEN**.

Slika D.11. prikazuje MAE radno okruženje nakon učitavanja jednog komentara.

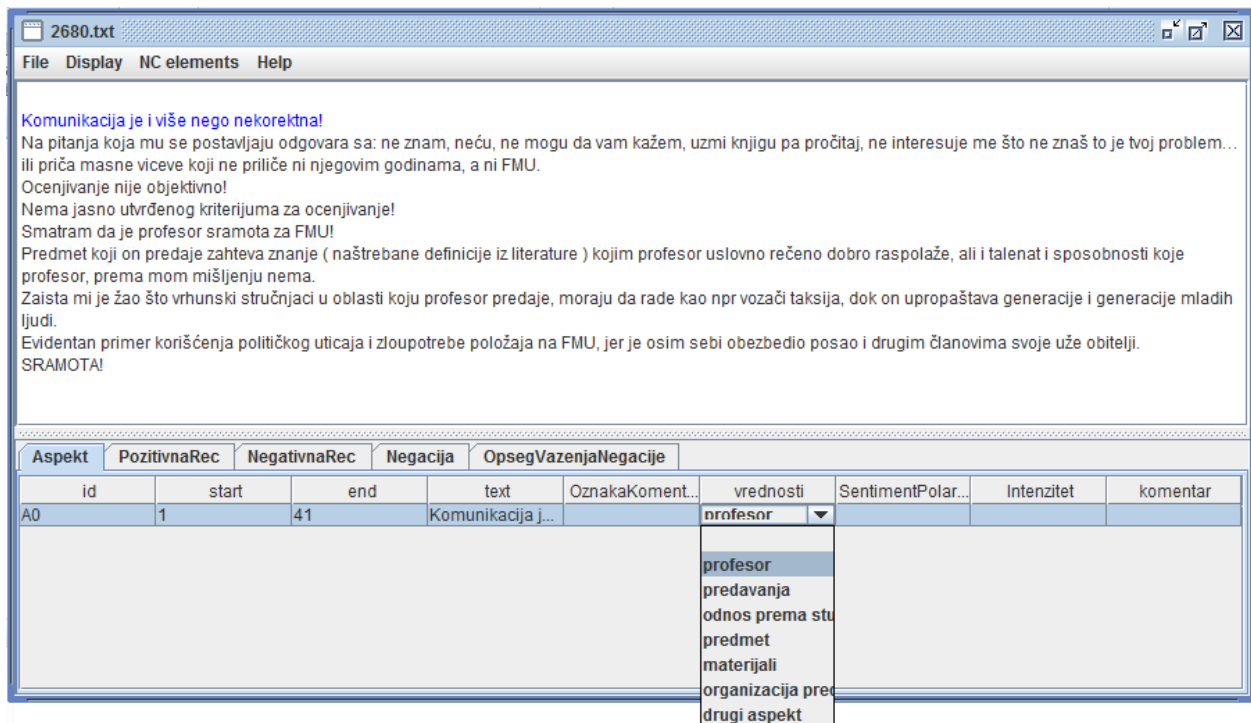


Slika D.11. Radno okruženje nakon učitavanja podataka

## D.7. Postupak anotacije

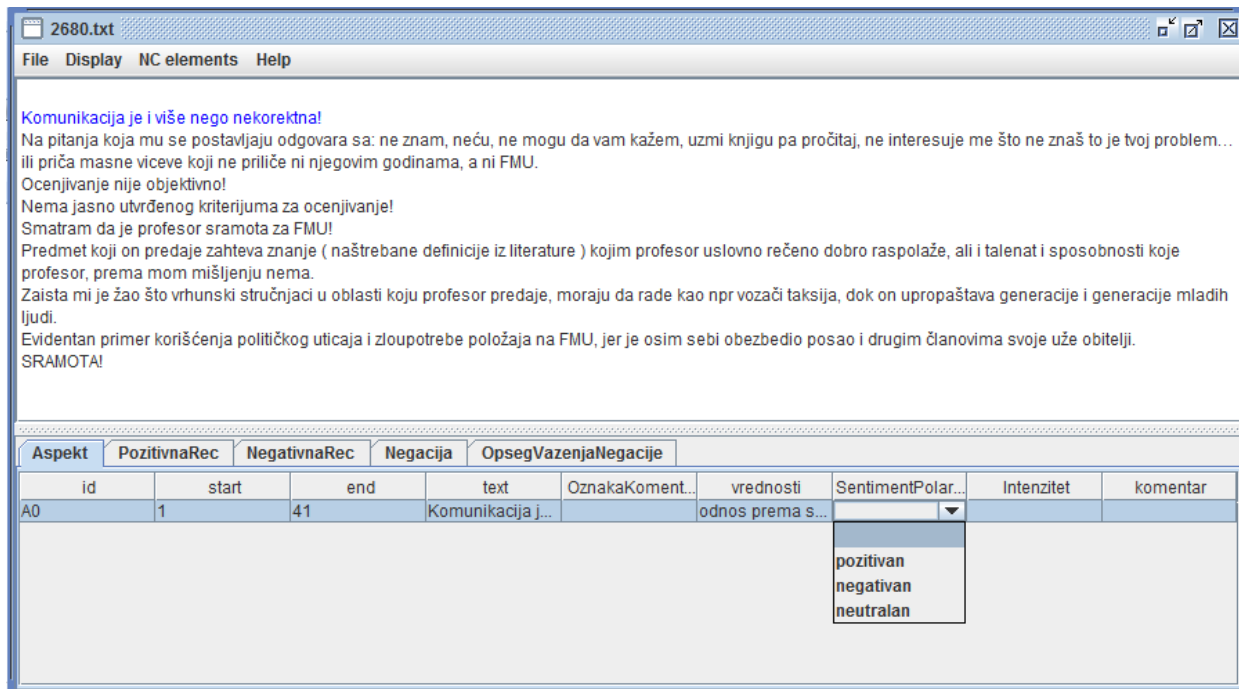
Nakon učitavanja DTD sheme i podataka nad kojima će anotator raditi anotaciju, može se pristupiti samom postupku anotacije. Naredne slike ilustruju korake anotacije.

Kako je već ilustrovano na slici D.1, proces anotacije započinje dodeljivanjem vrednosti atributu Aspekt. Jedan red predstavlja jednu rečenicu komentara. Potrebno je rečenicu pročitati i levim tasterom miša obuhvatiti deo teksta koji se anotira. Na selektovani deo teksta kliknuti desnim tasterom miša i sa padajućeg menija odabrati **ASPEKT**. Dodeljivanje vrednosti atributu aspekt je prikazano na slici D.12. Kako se na slici vidi, rečenica koja je predmet anotacije menja boju.



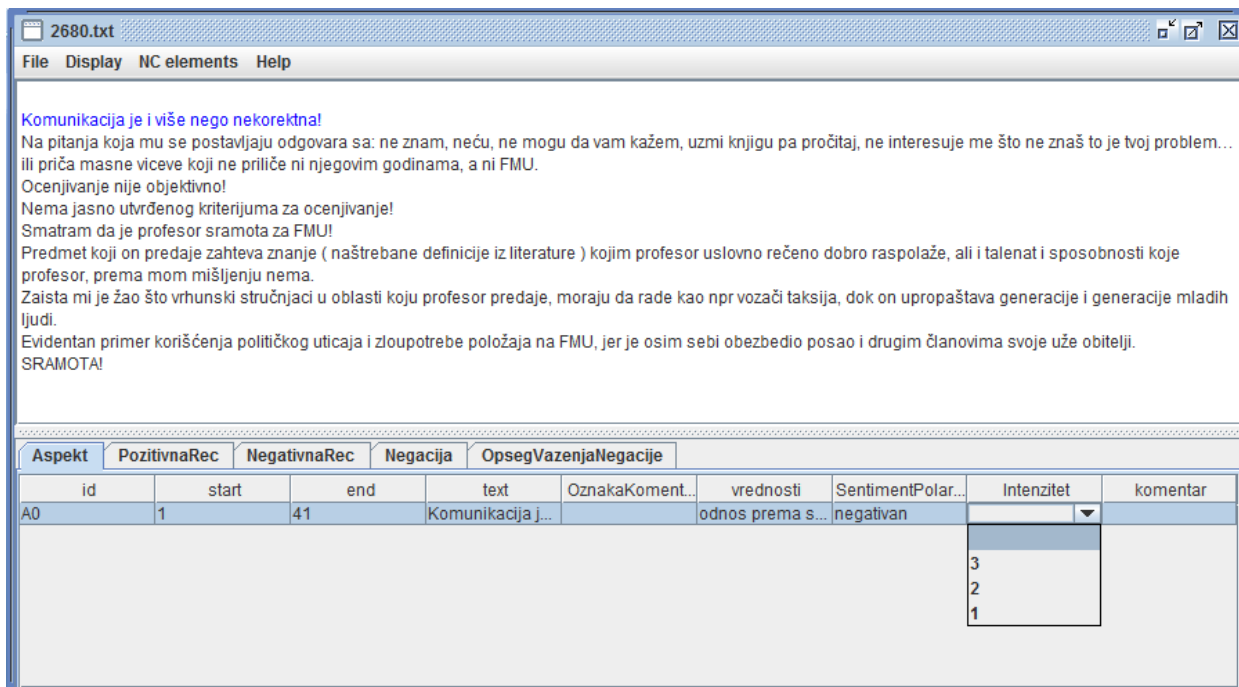
Slika D.12. Odabir vrednosti za atribut Aspekt

Sledeći korak anotacije je dodeljivanje sentimenta iskazanom aspektu, slika D.13.



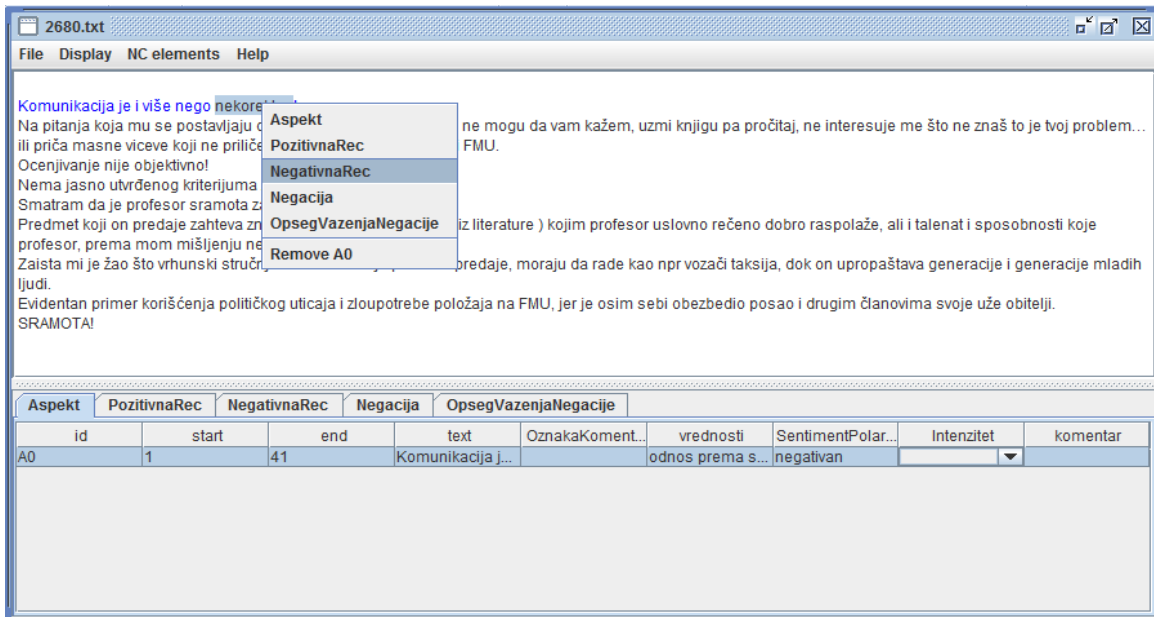
Slika D.13. Definisanje vrednosti atributa Sentiment Polaritet

Identifikovanom sentimentu je potrebno dodeliti intenzitet, slika D.14.



Slika D.14. Definisanje vrednosti atributa Sentiment intenzitet

U narednom koraku anotator identifikuje izraze sentimenta, selektuje ih u tekstu (jedan po jedan) i označava ih u zavisnosti od toga da li je reč o pozitivnom ili negativnom sentimentu, slika D.15.



Slika D.15. Označavanje sentiment izraza

Ukoliko anotator ima dodatni komentar u vezi sa anotacijom ili rečenicom, može ga upisati u polje **Komentar**. Ovo polje se prvenstveno koristi ukoliko je anotator nesiguran u svoju anotaciju, te je potrebno da navede stepen pouzdanja u tačnost anotacije.

MAE alat olakšava anotaciju rečenica u kojima je opisano više od jednog aspekta ili su iskazana osećanja različitog polariteta. Ukoliko je u rečenici prisutno više aspekata, selektovati samo deo rečenice koji se odnosi na jedan aspekt i anotirati ga, slika D.16. i D.17. Nakon toga anotirati sledeći deo rečenice, u kojem se govori o drugom aspektu.

2703.xml

File Display NC elements Help

Jedino što je dotična profesorka uradila jeste priprema ridera za kurseve koje vodi na osnovnim i doktorskim studijama muzikologije, kao i na kursevima uporedne istorije muzike na osnovnim studijama.  
 Predavanja, takoreći, ne drži, tera studente da čitaju tekstove i referišu obračujući pažnju na irelevantne činjenice.  
 Ne ume da preuzme ulogu moderatora.  
 Uvek sam imala utisak da ne zna o čemu se radi na njenim kursevima.  
 Zakida poene kako stigne i ume da bude i nekorektna.  
 Pritom, opterećuje studente faktografijom, prikupljanjem i bubanjem činjenica, bez ikakvog podsticaja za njihovom kontekstualizacijom.  
 Ipak, ne mogu reći da je kao čovek loša.  
 Jednostavno, nije dobra kao pedagoq.

Aspekt	PozitivnaRec	NegativnaRec	Negacija	OpsegVazenjaNegacije					
id	start	end	text	OznakaKoment...	vrednosti	SentimentPolar...	Intenzitet	komentar	
A0	1	201	Jedino što je d...	2703	profesor	negativan	1		
A1	202	232	Predavanja, tak...	2703	predavanja	negativan	2		
A2	233	320	tera studente d...	2703	odnos prema s...	negativan	2		
A3	321	356	Ne ume da pre...	2703	profesor	negativan	2		
A5	357	424	Uvek sam imal...	2703	profesor	negativan	3		
A6	425	477	Zakida poene k...	2703	odnos prema s...	negativan	3		
A7	478	612	Pritom, optereć...	2703	odnos prema s...	negativan	3		
A8	613	653	Ipak, ne mogu ...	2703	profesor	pozitivan	1	rec?	

Slika D.16. Anotacija rečenice u kojoj je opisano više aspekata

2703.xml

File Display NC elements Help

Jedino što je dotična profesorka uradila jeste priprema ridera za kurseve koje vodi na osnovnim i doktorskim studijama muzikologije, kao i na kursevima uporedne istorije muzike na osnovnim studijama.  
 Predavanja, takoreći, ne drži, tera studente da čitaju tekstove i referišu obračujući pažnju na irelevantne činjenice.  
 Ne ume da preuzme ulogu moderatora.  
 Uvek sam imala utisak da ne zna o čemu se radi na njenim kursevima.  
 Zakida poene kako stigne i ume da bude i nekorektna.  
 Pritom, opterećuje studente faktografijom, prikupljanjem i bubanjem činjenica, bez ikakvog podsticaja za njihovom kontekstualizacijom.  
 Ipak, ne mogu reći da je kao čovek loša.  
 Jednostavno, nije dobra kao pedagoq.

Aspekt	PozitivnaRec	NegativnaRec	Negacija	OpsegVazenjaNegacije					
id	start	end	text	OznakaKoment...	vrednosti	SentimentPolar...	Intenzitet	komentar	
A0	1	201	Jedino što je d...	2703	profesor	negativan	1		
A1	202	232	Predavanja, tak...	2703	predavanja	negativan	2		
A2	233	320	tera studente d...	2703	odnos prema s...	negativan	2		
A3	321	356	Ne ume da pre...	2703	profesor	negativan	2		
A5	357	424	Uvek sam imal...	2703	profesor	negativan	3		
A6	425	477	Zakida poene k...	2703	odnos prema s...	negativan	3		
A7	478	612	Pritom, optereć...	2703	odnos prema s...	negativan	3		
A8	613	653	Ipak, ne mogu ...	2703	profesor	pozitivan	1	rec?	

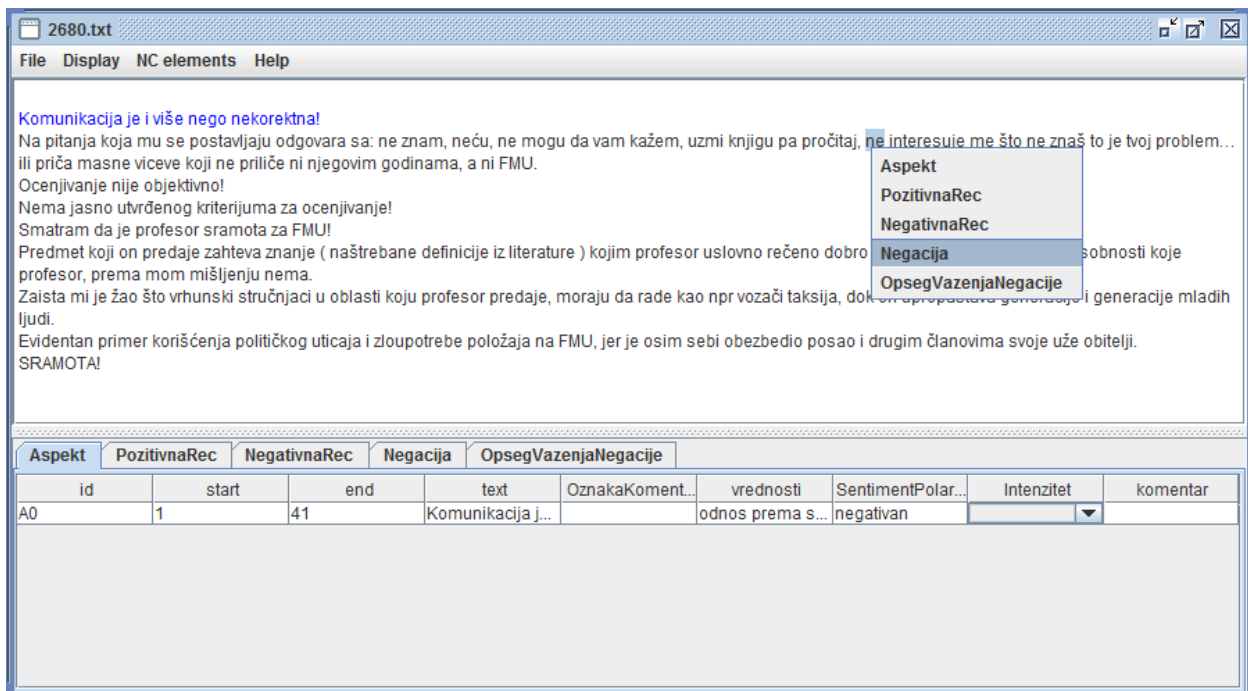
Slika D.17. Anotacija rečenice u kojoj je opisano više aspekata

Ukoliko je u jednoj rečenici pokriven jedan aspekt, ali je prisutno više različitih osećanja, takođe anotirati deo po deo rečenice, slika D.18.

Obrazložit ću svoje glasanje temeljno.  
 Predavanja nisu korisna ni u kakvom kontekstu.  
 Predavanja treba da se PREDAJU, a ne da se čita sa slajda sopstvena skripta.  
 Samim tim, druga tačka, odnosno stepen interesantnosti predavanja jednak je nuli.  
Nije zanimljivo čitati sa slajdova puke informacije.  
 što se tiče komunikacije, to je u redu, mada ni od toga nema koristi ier se onet sve svede na nočetnu tezu.  
Materijal za učenje jeste dostupan ali je nerazumljiv, nabacan, bez reda, bez nekog razumljivog rečnika.  
 Takođe, moram da dodam da se rešenja testova koje profesorka daje studentima ne poklapaju sa informacijama iz njene skripte.  
 Prema tome, ocena 1.75 je zaista zaslužena.

Slika D.18. Anotiranje jednog aspekta sa dva sentimenta

Anotiranje negacije započinje selektovanjem reči, odnosno rečice, koja ukazuje na negaciju. Desnim tasterom kliknuti na selektovanu reč/rečcu i sa padajućeg menija odabrati *Negacija*, slika D.19.



Slika D.19. Označavanje negacije

Opseg važenja negacije se prvo selektuje u tekstu komentara, potom se sa padajućeg menija odabira opcija *OpsegVazjenjaNegacije*, slika D.20.

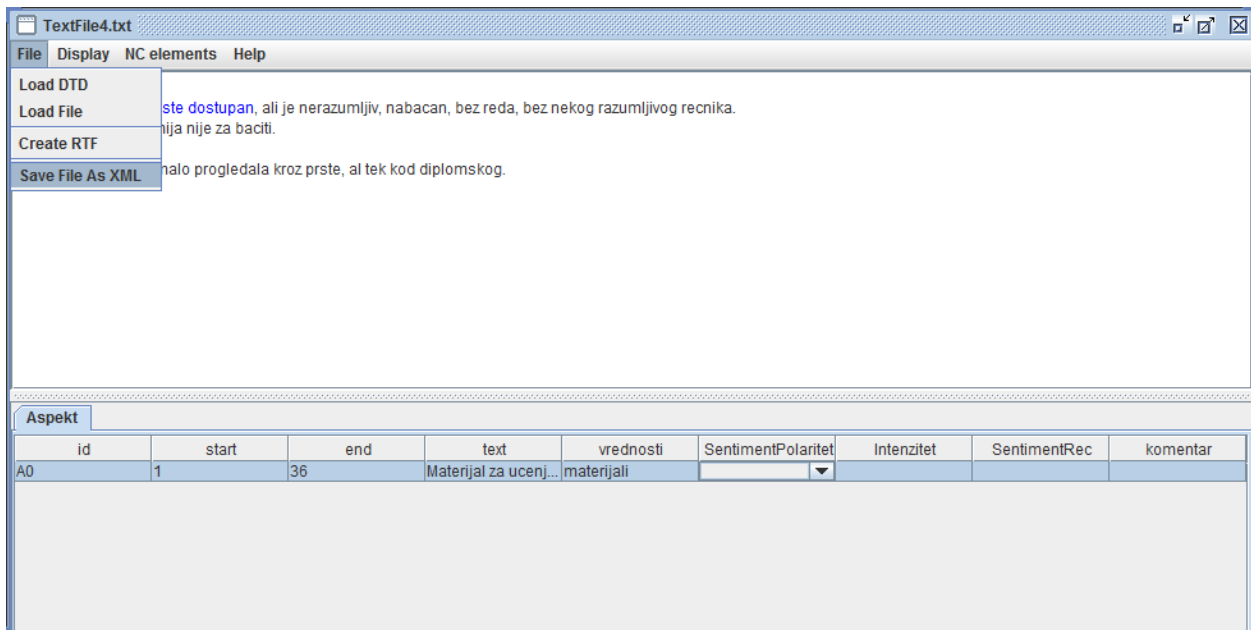
Komunikacija je i više nego nekorektna!  
 Na pitanja koja mu se postavljaju odgovara sa: ne znam, neću, ne mogu da vam kažem, uzmi knjigu pa pročitaj, ne interesuje me što ne znaš to je tvoj problem  
 Ili priča masne viceve koji ne priliče ni njegovim godinama, a ni FMU.  
 Ocenjivanje nije objektivno!  
 Nema jasno utvrđenog kriterijuma za ocenjivanje!  
 Smatram da je profesor sramota za FMU!  
 Predmet koji on predaje zahteva znanje ( naštrebane definicije iz literature ) kojim profesor uslovno rečeno dobro raspolaze, ali i talenat i sposobnosti  
 profesor, prema mom mišljenju nema.  
 Zaista mi je žao što vrhunski stručnjaci u oblasti koju profesor predaje, moraju da rade kao npr vozači taksija, dok on upropaštava generacije i generacije mladih ljudi.  
 Evidentan primer korišćenja političkog uticaja i zloupotrebe položaja na FMU, jer je osim sebi obezbedio posao i drugim članovima svoje uže obitelji.  
 SRAMOTA!

id	start	end	text	OznakaKoment...	vrednosti	SentimentPolar...	Intenzitet	komentar
A0	1	41	Komunikacija j...		odnos prema s...	negativan		

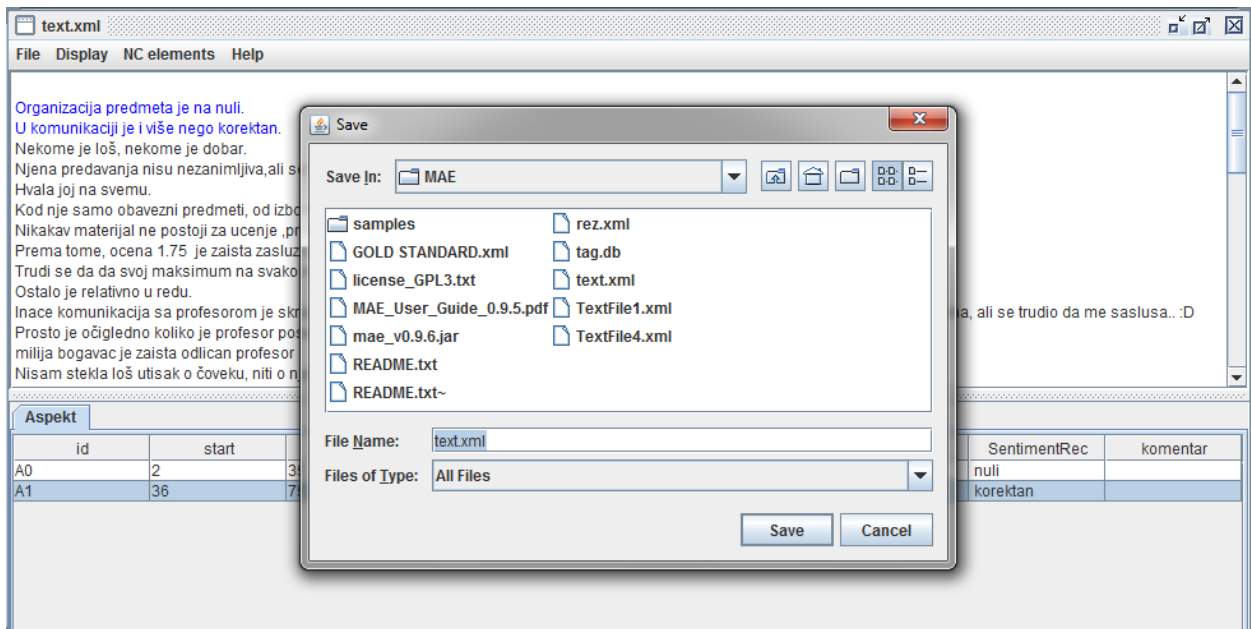
Slika D.20. Anotiranje opsega važenja negacije

## D.8. Čuvanje anotacije i nastavak rada

Kako bi se sačuvala informacije koje su anotirane, potrebno je iz menija **File** odabrati **Save as XML**, slike D.21. i D.22. Na ovaj način će se kreirati dokument koji sadrži tekst i oznake koje su anotatori kreirali.



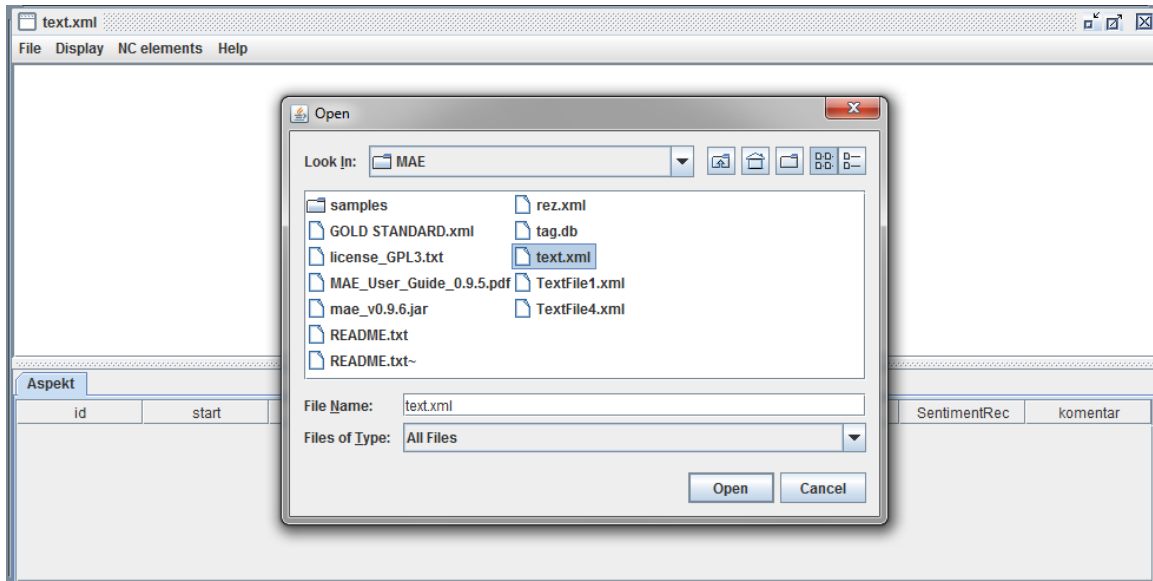
Slika D.21. Čuvanje anotiranih informacija 1. korak



Slika D.22. Čuvanje anotiranih informacija 2. korak

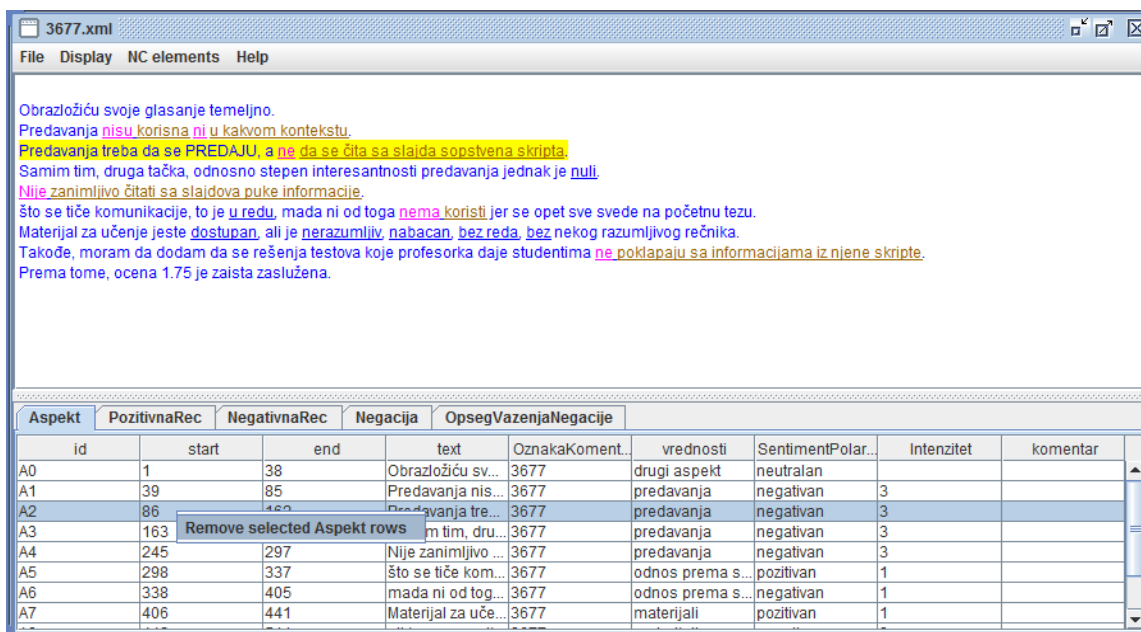


Da bi se nastavio rad, potrebno je ponovo učitati fajl dtdSchema iz **File** menija i nakon toga sa istog menija potrebno je odabrati **Load File** opciju. Ukoliko se vrše korekcije u anotaciji, umesto tekstualnog fajla učitati xml fajl sačuvan u prethodnom koraku, slika D.23. Ukoliko se radi nova anotacija, učitati neanotirani tekstualni fajl. Nakon konačne anotacije svaki tekstualni fajl mora da ima prateći xml fajl u kojem se nalaze anotirane informacije.



Slika D.23. Učitavanje anotiranih informacija i nastavak rada

Ukoliko nastane greška u postupku anotacije, na primer obuhvaćeno je više ili manje teksta od željenog, potrebno je u donjoj polovini ekrana kliknuti desnim tasterom miša na red sa pogrešnom anotacijom i odabrati opciju **Remove selected Aspect rows**, čime će ta konkretna anotacija biti uklonjena sa liste i iz fajla i omogućiti korigovanje anotacije, slika D.24.



Slika D.24. Uklanjanje pogrešne anotacije

## E. Procena stepena saglasnosti anotatora

Prilozi ilustruju formu podataka nad kojima se vršila procena stepena saglasnosti anotatora po pojedinim atributima. Slika E.1. ilustruje anotacije po anotatorima, gde su crvenom bojom označena odstupanja u anotacijama.

	Anotator 1			Anotator 2			Anotator 3			Anotator 4		
Izuzetno dobar predavač, harizmatičan i duhovit čovek,	profesor	pozitivan	3	profesor	pozitivan	3	profesor	pozitivan	3	Profesor	Pozitivan	3
ali nikakav pedagog, ume da omalovažava studente i da bude ciničan.	odnos prema studentima	negativan	3	odnos prema studentima	negativan	2	odnos prema studentima	negativan	3	Odnos prema studentima	Negativan	3
Komunikativan, elokventan, jasan i dovoljno glasan.	profesor	pozitivan	3	profesor	pozitivan	2	profesor	pozitivan	3	Profesor	Pozitivan	3
Komunikativan, predusretljiv, ažuran.	profesor	pozitivan	3	profesor	pozitivan	2	profesor	pozitivan	2	Profesor	Pozitivan	3
Korektan, izuzetno stručan.	profesor	pozitivan	3	profesor	pozitivan	2	profesor	pozitivan	3	Profesor	Pozitivan	3
Ma dobar je,	profesor	pozitivan	1	profesor	pozitivan	2	profesor	pozitivan	2	Profesor	Pozitivan	2
samo je onako plašljiv, znate.	profesor	negativan	1	profesor	negativan	1	profesor	negativan	2	Profesor	Negativan	1
Zna on svašta,	profesor	pozitivan	2	profesor	pozitivan	2	profesor	pozitivan	2	Profesor	Pozitivan	2
ali samo da je malo više muževan.	profesor	neutralan	0	drugi aspekt	negativan	2	profesor	negativan	2	Profesor	Negativan	1
Meni je držao vežbe iz predmeta o kojem on konkretno ništa ne zna , mislim da je greška što on drži vežbe iz tog predmeta	profesor	negativan	3	profesor	negativan	2	profesor	negativan	3	Profesor	Negativan	2
a inače je korektan.	profesor	pozitivan	2	profesor	pozitivan	2	profesor	pozitivan	2	Profesor	Pozitivan	2
Možda je to posledica toga što sam ga dve godine bombardovao mail-ovima, ali uvek je bio nekako zvaničan i nepristupačan.	odnos prema studentima	negativan	2	odnos prema studentima	negativan	2	odnos prema studentima	negativan	2	Odnos prema studentima	Negativan	1
Nadmen, neprijatan i nekorektan.	profesor	negativan	3	profesor	negativan	2	odnos prema studentima	negativan	3	Profesor	Negativan	3
Neprofesionalan i nestručan.	profesor	negativan	3	profesor	negativan	2	profesor	negativan	3	Profesor	Negativan	3
Nije strog profesor	profesor	pozitivan	1	profesor	pozitivan	2	profesor	pozitivan	2	Profesor	Pozitivan	2
ali je dosadan.	profesor	negativan	1	profesor	negativan	2	profesor	negativan	2	Profesor	Negativan	3
Odličan asistent, jako dobro objašnjava, drži pažnju, nije nadmen i dosadan.	profesor	pozitivan	3	profesor	pozitivan	3	profesor	pozitivan	3	Profesor	Pozitivan	3
Odnos u komunikaciji sa studentima ja često vrlo snishodljiv i neprofesionalan.	odnos prema studentima	negativan	3	odnos prema studentima	negativan	2	odnos prema studentima	negativan	3	Odnos prema studentima	Negativan	3
Predavanja baš i nismo imali toliko,	predavanja	negativan	1	organizacija predmeta	negativan	2	predavanja	negativan	2	Organizacija predmeta	Negativan	1

Slika E1. Ilustrativni prikaz paralelnih anotacija po anotatorima

Slike E.2–E.4. ilustruju postupak kalkulacije stepena saglasnosti anotatora za attribute Sentiment polaritet, Aspekt i Intenzitet sentimenta, formu input i output podataka.

	A	B	C	D	E	F	G	H	I	J
1										
2			Pozitivan (j=1)	Negativan (j=2)	Neutralan (j=3)	$P_i$				
3		1	0	4	0	1.00			N	568
4		2	0	4	0	1.00			n	4
5		3	0	4	0	1.00			k	3
6		4	4	0	0	1.00				
7		5	4	0	0	1.00			$\bar{P}$	0.9830
8		6	0	4	0	1.00			$\bar{P}_e$	0.5253
9		7	4	0	0	1.00				
10		8	4	0	0	1.00			kappa	0.964
11		9	4	0	0	1.00				
12		10	0	4	0	1.00			Var( $\kappa$ )	0.001258
13		11	4	0	0	1.00			SE( $\kappa$ )	0.035
14		12	4	0	0	1.00				
15		13	4	0	0	1.00				
565		563	4	0	0	1.00				
566		564	4	0	0	1.00				
567		565	4	0	0	1.00				
568		566	4	0	0	1.00				
569		567	4	0	0	1.00				
570		568	4	0	0	1.00				
571		Ukupno	1400	867	5					
572		$P_j$	0.616	0.382	0.002					

Slika E.2. Računanje saglasnosti anotacije – Sentiment polaritet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1															
2			Profesor (j=1)	Predavanja (j=2)	Odnos prema studentima (j=3)	Predmet (j=4)	Materijali (j=5)	Organizacija predmeta (j=6)	Drugi aspekt (j=7)	$P_i$					
3		1	4	0	0	0	0	0	0	1.00				N	568
4		2	4	0	0	0	0	0	0	1.00				n	4
5		3	4	0	0	0	0	0	0	1.00				k	7
6		4	4	0	0	0	0	0	0	1.00					
7		5	4	0	0	0	0	0	0	1.00				$\bar{P}$	0.9149
8		6	0	0	4	0	0	0	0	1.00				$\bar{P}_e$	0.5907
9		7	4	0	0	0	0	0	0	1.00					
10		8	4	0	0	0	0	0	0	1.00				kappa	0.792
11		9	4	0	0	0	0	0	0	1.00					
12		10	4	0	0	0	0	0	0	1.00				Var( $\kappa$ )	0.002663
13		11	4	0	0	0	0	0	0	1.00				SE( $\kappa$ )	0.052
14		12	4	0	0	0	0	0	0	1.00					
15		13	4	0	0	0	0	0	0	1.00					
566		564	4	0	0	0	0	0	0	1.00					
567		565	2	0	2	0	0	0	0	0.33					
568		566	0	0	4	0	0	0	0	1.00					
569		567	4	0	0	0	0	0	0	1.00					
570		568	4	0	0	0	0	0	0	1.00					
571		Ukupno	1709	212	287	21	21	9	13						
572		$P_j$	0.752	0.093	0.126	0.009	0.009	0.004	0.006						

Slika E.3. Računanje saglasnosti anotacije – Aspekt

Procena stepena saglasnosti anotatora

	A	B	C	D	E	F	G	H	I	J
1										
2			<b>Jak (j=1)</b>	<b>Srednji (j=2)</b>	<b>Slab (j=3)</b>	<b>P<sub>i</sub></b>				
3		1	4	0	0	1.00				
4		2	4	0	0	1.00		N	563	
5		3	4	0	0	1.00		n	4	
6		4	4	0	0	1.00		k	3	
7		5	3	1	0	0.50				
8		6	2	2	0	0.33		$\bar{P}$	0.7487	
9		7	4	0	0	1.00		$\bar{P}_e$	0.5487	
10		8	3	1	0	0.50				
11		9	4	0	0	1.00		kappa	0.443	
12		10	2	0	2	0.33				
13		11	4	0	0	1.00		Var( <sub>i</sub> )	0.001756	
14		12	4	0	0	1.00		SE( <sub>i</sub> )	0.042	
15		13	4	0	0	1.00				
558		561	4	0	0	1.00				
559		562	4	0	0	1.00				
560		563	3	1	0	0.50				
561		564	4	0	0	1.00				
562		565	2	2	0	0.33				
563		566	4	0	0	1.00				
564		567	3	1	0	0.50				
565		568	0	4	0	1.00				
566		<b>Ukupno</b>	1563	571	118					
567		<b>P<sub>j</sub></b>	0.694	0.254	0.052					

Slika E4. Računanje saglasnosti anotacije – Sentiment intenzitet



### Merenje saglasnosti anotacija na nivou izraza

Anotacija na nivou izraza rađena je za attribute PozitivnaReč, NegativnaReč, Negacija i Opseg važenja negacije. Slika E.6. prikazuje odnos identifikovanih vrednosti po atributima, kao i broj anotacija koje su u saglasnosti. Može se primetiti da je identifikovano više pozitivnih sentiment reči u odnosu na negativne. Vrednosti iz prikazanih matrica predstavljaju inpute za kalkulisanje preciznosti, opoziva i F-mere.

	Pozitivna reč			
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
A <sub>1</sub>	516	457	411	436
A <sub>2</sub>	457	493	400	410
A <sub>3</sub>	411	400	427	385
A <sub>4</sub>	436	410	385	470

	Negacija			
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
A <sub>1</sub>	89	80	76	79
A <sub>2</sub>	80	81	71	73
A <sub>3</sub>	76	71	76	73
A <sub>4</sub>	79	73	73	80

	Negativna reč			
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
A <sub>1</sub>	315	243	242	226
A <sub>2</sub>	243	322	223	202
A <sub>3</sub>	242	223	255	203
A <sub>4</sub>	226	202	203	272

	Opseg važenja negacije			
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
A <sub>1</sub>	88	73	71	72
A <sub>2</sub>	73	81	63	67
A <sub>3</sub>	71	63	76	68
A <sub>4</sub>	72	67	68	80

Slika E.6. Matrice identifikovanog broja sentiment reči, odnosno negacije i njenog opsega po anotatorima

## F. Sentiment reči identifikovane u korpusu

U prilogu je data lista sentiment reči identifikovanih u fazi anotacije korpusa. Liste daju prikaz jedinstvenih oblika pozitivnih i negativnih reči identifikovanih u skupu za obučavanje i skupu za testiranje.

### F.1. Jedinstveni oblici pozitivnih sentiment reči u skupu za obučavanje

1. adekvatna	318. konstantnim	635. odlika	952. šećer
2. afekcije	319. konsultacija	636. odmeren	953. šef
3. akademski	320. ko on	637. odskaču	954. sensei
4. akademskih	321. korektnih	638. oduševila	955. setiti
5. akademskom	322. korektnijih	639. oduševila	956. shvata
6. aktuelna	323. korektno	640. oduševljena	957. shvate
7. aktuelne	324. korektno	641. odvojiš	958. shvati
8. ambiciozna	325. korišćenje	642. ogromnu	959. shvatim
9. analizira	326. korisnijim	643. ohrabri	960. shvatio
10. angažovana	327. korisnim	644. ohrabruje	961. shvatite
11. angažovanje	328. korisnost	645. ok	962. shvatiti
12. angažuje	329. koristi	646. ok	963. siguran
13. angažuju	330. koristiti	647. olakšavajuće	964. sigurno
14. animira	331. kostu vujića	648. oličenje	965. simpatičan
15. as	332. kralj	649. oličenje	966. širi
16. autoritativan	333. kraljevima	650. omiljen	967. širok
17. autoritet	334. kraljica	651. omogućuje	968. širokim
18. ažuran	335. kraljina	652. opravdan	969. sjajna
19. ažuran	336. krasni	653. opseg	970. sjajno
20. ažurna	337. kreativan	654. opšte	971. sjajnom
21. baca te u vatru	338. kreativna	655. opušten	972. skidam kapu
22. bavi	339. kreativne	656. opuštena	973. skladu
23. baza	340. kreativnom	657. opuštenijom	974. školovan
24. beskonačan	341. kreativnošću	658. opušteno	975. slatkiš
25. besprekora	342. kul	659. opusti	976. sluša
26. bez greške	343. kultura	660. organizovan	977. slušajte
27. bez ikakve nadmenosti	344. kulturom	661. organizovanu	978. slušaš
28. bez ikakvih	345. kvalitetan	662. oslobađajući	979. slušati
29. bez ikakvih problema	346. kvalitete	663. oslobodi	980. šmeker
30. bez komentara	347. kvalitetne	664. oslobodim	981. smernica
31. bez previše	348. kvalitetni	665. osmehom	982. smešan
32. bezrezervna	349. kvalitetnih	666. osmišljena	983. smiren

33. bez stresa	350. kvalitetnim	667. osnova	984. smirena
34. bez zamerki	351. kvalitetno	668. osnovno	985. smirenom
35. bilo kad	352. kvalitetu	669. ostaje	986. smislu
36. bilo šta	353. laf	670. ostvari	987. snalaženje
37. biser	354. lafica	671. otkloniti	988. snalažljivost
38. bisera	355. lagana	672. otvoren	989. solidan
39. bistrog	356. laka	673. ozbiljan	990. solidna
40. bitan	357. lakši	674. ozbiljnim	991. solidna
41. bitne	358. lav	675. pametan	992. solidno
42. bitnim	359. legenda	676. pametne	993. šoumen
43. blag	360. legendo	677. pametnih	994. specifičan
44. blizak	361. lep	678. pametniji	995. spektar
45. blizak	362. lepa	679. pametnijih	996. spontano
46. bodri	363. lepe	680. pametnim	997. spontanost
47. boem	364. lepi	681. pamte	998. sposobnošću
48. bog	365. lepog	682. pamtim	999. sposobnosti
49. bog	366. lepom	683. pamtim	1000. spreman
50. bogata	367. lepotom	684. pamtiti	1001. srčan
51. bogatim	368. lepu	685. parira	1002. sreće
52. bogato	369. lik	686. pažnjom	1003. srećni
53. bogatstvo	370. lik	687. pedagoga	1004. srećnih okolnosti
54. boginja	371. likčina	688. pedagoški	1005. sreću
55. bolje	372. likčina	689. pedantan	1006. staložnost
56. boljeg	373. ljubazna	690. perfektan	1007. stara garda
57. boljim	374. ljudina	691. perfektna	1008. stara škola
58. bori	375. ljudina	692. perfektno	1009. stečenom
59. bravo	376. ljudinom	693. perspektivan	1010. stekne
60. briljantnog	377. ljudski	694. perspektivan	1011. šteta
61. brižna	378. logična	695. perspektivom	1012. šteta
62. brz	379. logično	696. pet	1013. stiču
63. brza	380. majka	697. pismen	1014. strog
64. budnim	381. majstorima	698. plemeniti	1015. stroga
65. budućnost	382. maksimalna	699. plus	1016. strpljenje
66. car	383. maksimalni	700. počastvošana	1017. stručan
67. carica	384. malobrojnih	701. počastvovani	1018. stručan
68. carina	385. malo vremena	702. počastvovanim	1019. stručna
69. ćaskanje	386. maniri	703. podjednako	1020. stručnim
70. čast	387. manirima	704. podlogu	1021. stručnjak
71. časti	388. maštovita	705. podršku	1022. stručnjaka
72. cenjen	389. maštovito	706. podržava	1023. stručno
73. cenjene	390. medalju	707. podsećaju	1024. stručnošću
74. cenjenoj	391. melem	708. podsticajna	1025. stručnost



75. centar	392. menja život.	709. podsticalo	1026. stručna
76. čini	393. merilima	710. podsticanje	1027. studije slučaja
77. čista	394. merodavan	711. podsticanjem	1028. stvaralačku
78. čista	395. metodičar	712. podsticao	1029. sujetan
79. čudo	396. metodički	713. podstiču	1030. super
80. čuješ	397. metodološki	714. podstrek	1031. svaka čast
81. čuvam	398. miran	715. pogled na život menja	1032. svašta
82. čuvati	399. mirna	716. pohvale	1033. sveobuhvatna
83. :-d	400. mjau	717. pohvali	1034. sve od sebe
84. :d	401. mlada	718. pohvalila	1035. svesna
85. daje	402. mnogo	719. pohvaliti	1036. svestran
86. daleko ćete dogurati	403. mnogo	720. po imenu	1037. svestrana
87. dambldor	404. mnogo brojih	721. pojašnjava	1038. svetla tačka
88. dati	405. mnogo više	722. pojedinačno	1039. svetlih
89. delija	406. model	723. poklanja	1040. svetska
90. detaljan	407. moderna	724. pokloni	1041. svetski
91. dijapazon	408. modernog	725. pokretačka	1042. svežine
92. dika	409. moguć	726. položi	1043. svideo
93. dinamična	410. moguće	727. položiceš	1044. svojim rečima
94. dinamična	411. moralnim	728. položicete	1045. tačan
95. dinamično	412. moralno	729. položila	1046. tačan
96. direktan	413. motivisao	730. položili	1047. tačno
97. direktna	414. motiviše	731. položim	1048. talenta
98. direktno	415. motiviše	732. položio	1049. tatko
99. disciplinovana	416. mozak	733. položiš	1050. temeljan
100. disciplinu	417. multidisciplinarna	734. položite	1051. temeljit
101. divan	418. nada	735. pomažu	1052. temeljna
102. divim	419. nada	736. pomogla	1053. top
103. divljenje	420. nadahnuće	737. pomogle	1054. topao
104. divnom	421. nadahnuta	738. po muke	1055. toplá
105. dobije	422. nađe	739. ponos	1056. tradicionalnog
106. dobra	423. nadu	740. ponos	1057. trag
107. dobro	424. nagrade	741. ponosna	1058. trajno
108. dobrom	425. nagrađen	742. ponudi	1059. traži
109. dobronameran	426. nagradice	743. popravila	1060. treba
110. dobronamerna	427. na graditi	744. posavetuje	1061. trebati
111. dobronamerni	428. naj	745. posebne	1062. truda
112. dodatna	429. najaktuelnijim	746. posećena	1063. trudio
113. dodatno	430. najbitnijem	747. posećivao	1064. tu
114. dođe	431. najbitniji	748. posetiti	1065. tutor
115. dodir	432. najbogatiji	749. pošten	1066. tutora
116. dogovora	433. najbolje	750. poštena	1067. tu za nas

117. dokazala	434. najboljeg	751. pošteni	1068. učenje
118. dokopate	435. najbolji	752. pošteno	1069. učešće
119. domaćin	436. najboljim	753. postignuća	1070. učini
120. dopada	437. najboljnih	754. poštovanje	1071. učinila
121. dopadljiv	438. najdetalnija	755. poštovanjem	1072. učite
122. dopao	439. najdivnija	756. poštovati	1073. učitelj
123. dosadan	440. najdivniji	757. poštujem	1074. učiti
124. dosledna	441. najdostojniji	758. posvećen	1075. učtiv
125. dostavi	442. najduhovitiji	759. posvećena	1076. ugledale
126. dostojni	443. najduže	760. posvećeno	1077. ugledam
127. dostupna	444. najharizmatičnijih	761. posvećenog	1078. ukapira
128. dostupnost	445. najinteresantnija	762. posvećenost	1079. ukaže
129. drag	446. najinteresantnija	763. posvećuje	1080. ukazuje
130. drago	447. najinteresantnijih	764. potencijal	1081. uključuje
131. dragocenost	448. najjača	765. potkrepljen	1082. uključen
132. dragulj	449. najjači	766. potkrepljena	1083. uključeni
133. draže	450. najjači	767. potpuniji	1084. uključiti
134. drug	451. najkompetitivniji	768. potpunosti	1085. uključuje
135. drugačija	452. najkompleksnije	769. potreban	1086. ulaže
136. drugačije	453. najkompletniji	770. potrebno	1087. ulepša
137. drugar	454. najkorektniji	771. poveća	1088. ulepšali
138. društvena	455. najkorektnija	772. povereni	1089. u letu
139. druželjubiva	456. najkorektnije	773. povezivanje	1090. um
140. druženje	457. najkorektniji	774. povezuje	1091. ume
141. drži pažnju	458. najkorektniji	775. povezuje mo	1092. umeću
142. držite	459. najkorektnijih	776. povučena	1093. umeju
143. dubinske	460. najkorektnijim	777. poželeo	1094. umeren
144. duge	461. najkorisnijim	778. poželjno	1095. u minut
145. duhovit	462. najkreativnija	779. pozitivne	1096. unapredi
146. duhovita	463. najkreativnija	780. poznavanje	1097. unapređivanje
147. dušu	464. najkvalitetnije	781. požrtvovan	1098. upornosti
148. edukativna	465. najkvalitetniji	782. praksi	1099. upotpunjavali
149. edukativnim	466. najbolji	783. praksom	1100. upozna
150. efikasna	467. najbolji	784. praktičan	1101. upućen
151. ekspert	468. najljubaznijih	785. praktičar	1102. upućenost
152. ekspresno	469. najneposredniji	786. praktična	1103. upućuje
153. ekstra	470. najobrazovanijih	787. praktičnog	1104. uravnoteženih
154. elanom	471. najorganizovaniji	788. praktičnost	1105. uređeniji
155. emeritusa	472. najorginalnijih	789. prati	1106. usavršava
156. enciklopedija	473. najosnovnije	790. pravedan	1107. usavršavanja
157. enciklopediju	474. najposećenija	791. pravedan	1108. usavršavanje
158. energičan	475. najposećenija	792. pravedna	1109. u službi
159. energičnost	476. najpošteniji	793. pravi	1110. uslužna

160. energija	477. najpozitivnija	794. pravično	1111. uslužnost
161. energiju	478. najpozitivnijem	795. pravilnog	1112. usmeri
162. enormno	479. najprijatnija	796. pravo	1113. usmeriti
163. entuzijasta	480. najprijatniji	797. pravog	1114. uspeh
164. entuzijastičan	481. najpristupačnijih	798. pravom smislu	1115. uspehom
165. entuzijizam	482. najrazličitijih	799. precizan	1116. uspehu
166. entuzijazma	483. najrealnije	800. precizna	1117. uspela
167. entuzijazmom	484. najrealniji	801. precizniji	1118. uspela
168. etičnosti	485. najsadržajnijim	802. predan	1119. uspeli
169. fantastičan	486. najširom	803. predana	1120. uspešan
170. fantastičan	487. najsitnijeg	804. predana	1121. uspešno
171. fantastična	488. najslađa	805. predanošću	1122. uspevala
172. fantastično	489. najslađe	806. predavao	1123. uštedi
173. fantastičnoj	490. najsposobnijih	807. predivna	1124. usvajanje
174. fascinira	491. najsrdanija	808. predobar	1125. usvojim
175. fenomenalan	492. najstručniji	809. predstavi	1126. utisak
176. fenomenalna	493. najvažnije	810. predstavlja	1127. u toku
177. fenomenalna	494. najveća	811. predstavnik	1128. uvek tu
178. fenomenalne	495. najveća	812. predusretljiva	1129. uveo
179. fenomenalno	496. najveće	813. predusretljivih	1130. uvodi
180. fer	497. najveći	814. predusretljivosti	1131. užas
181. fer	498. najveći	815. preduzimljiva	1132. uživala
182. fin	499. najviše	816. pregleda	1133. uživam
183. fina	500. najviše	817. pregledne	1134. uživam
184. finom	501. najvišem	818. prelep	1135. uživanje
185. fleksibilni	502. najvišim	819. preljubazan	1136. uznapredujem
186. fokusirana	503. najzabavnijih	820. prenesu	1137. uz nas
187. fora	504. najzanimljivija	821. preneti	1138. uzor
188. frajer	505. najzanimljivija	822. prenetom	1139. valja
189. garantujem	506. najzanimljiviji	823. prenosi	1140. vanserijski
190. genijalac	507. najzanimljiviji	824. preporučujem	1141. vaspitavao
191. genijalac	508. naklon	825. preporuku	1142. važan
192. genijalan	509. napredovalo	826. prepozna	1143. važni
193. genijalan	510. napretka	827. prepun	1144. važnih
194. genijalnom	511. naš	828. prepune	1145. važnijim
195. genije	512. našali	829. presladak	1146. važnost
196. glasan	513. nasmejani	830. presrećni	1147. većem
197. glavni	514. nasmejana	831. prestala	1148. veći
198. gospođa	515. nasmejani	832. prevaziđem	1149. većim
199. gospodski	516. natprosečno	833. previše	1150. vedra
200. gospoja	517. naučio	834. prezadovoljan	1151. veličini
201. gotivan	518. naučite	835. prezadovoljna	1152. velike
202. gotivan	519. naučna	836. prihvatljivo	1153. veliki

203. gotivili	520. navede	837. prijatelja	1154. veliko
204. gotivna	521. na zemlji	838. prijateljski	1155. veruje
205. govornik	522. ne bi uspelo	839. prijatna	1156. vesela
206. harizma	523. ne bubanje	840. prijatno	1157. veselo
207. harizmatična	524. neće imati problema	841. prijatnu	1158. vešt
208. harizme	525. neće mnogo	842. prikazuju	1159. vešti
209. harizmu	526. nećemo imati	843. prilagodi	1160. veštine
210. humor	527. neću promeniti	844. priljatelj	1161. vežba
211. humora	528. ne dirajte	845. primenjeno	1162. vice ve
212. hvala	529. neformalna	846. primenjivog	1163. vickasta
213. hvaliti	530. ne govati	847. primenjuju	1164. video
214. idealno	531. ne guje	848. primenljiva	1165. vijuge
215. idealnog	532. neiskompleksirana	849. pripremi	1166. više mladih
216. ide uz	533. neizbrisiv	850. pristaje	1167. viši
217. idite	534. ne ljute	851. pristupačan	1168. visoke
218. ijedne zamerke	535. nema bolje	852. pristupačna	1169. visokim
219. ilustrativno	536. nema dalje	853. pristupačno	1170. višu
220. imaju uvide	537. nema digresija	854. pristupačnost	1171. vizionar
221. ima vremena	538. nema dosadnih	855. prisutno	1172. vlada
222. impozantnog	539. nema omiljene	856. privlačenja	1173. vodi
223. impresivna	540. nema poklanjanja	857. prizemna	1174. vođstvo
224. ineteresantna	541. nemaš komentara	858. prodiskutuje	1175. vokabular
225. informacija	542. nema te barijere	859. prodoran	1176. vole
226. inovativan	543. nemerivog	860. produktivnog	1177. volela
227. inovativna	544. nemerljivo	861. profesionalac	1178. voli
228. inovativna	545. ne može bolje	862. profesionalan	1179. voljan
229. inovativnošću	546. ne okleva	863. profesionalan	1180. voljeni
230. inspirativan	547. ne omalovažava	864. profesionalizmu	1181. volju
231. inspirativna	548. ne pogleda	865. profesionalni	1182. vrati
232. inspirišu	549. neposredna	866. profesionalno	1183. vratite
233. intelektualna	550. neposrednost	867. profesionalnost	1184. vredan
234. intelektualnom	551. nepravedno	868. profesionalnost	1185. vrede
235. interakcijom	552. ne prekida	869. profesorčina	1186. vredeo
236. interakciju	553. ne preopterećuje	870. progledala	1187. vrednih
237. interaktivan	554. nepresušan izvor	871. prolaze	1188. vrednim
238. interaktivna	555. neprevaziđen	872. proleti	1189. vrednost
239. interaktivno	556. neprocenjivo	873. promene	1190. vrhunski
240. interdisciplinaran	557. neprolazne	874. promovisanju	1191. vrhunski
241. interesantne	558. ne rađa dva puta	875. propustio	1192. vrlina
242. interesantnija	559. nesebično	876. prosečna	1193. vrsni
243. interesantnijih	560. ne smara	877. proširio	1194. vrzma
244. interesantnijim	561. nestrljenjem	878. provežbati	1195. zabavan
245. interesantno	562. nesvakidašnji	879. pruža	1196. zabavna

246. interesovanja	563. ne tera	880. pružili	1197. zadovoljavajuća
247. interesovanjem	564. ne trepćem	881. prvak	1198. zadovoljna
248. intuitivna	565. neumoran	882. prvom	1199. zadovoljstvo
249. inventivnost	566. neumorna	883. pun	1200. zadovoljstvom
250. ironičan	567. neuporedivo	884. pusti	1201. zadrži
251. iščekujete	568. neverovatan	885. rad	1202. zagantovan
252. iscrpno	569. neverovatna	886. radiš	1203. zahtevan
253. iskazuje	570. neviđeni	887. raj	1204. zahvalim
254. iskoristila	571. ne voli	888. ranije	1205. zahvalna
255. iskren	572. ne zakida	889. raspoláže	1206. zainteresovana
256. iskreno	573. ne zaostaju	890. razgovara	1207. zainteresovanim
257. iskrenu	574. nezaustavljiv	891. razlikuje	1208. zalaganje
258. iskusan	575. nije agresivna	892. razložnog	1209. zalaganjem
259. iskustvene	576. nije cepidlaka	893. razmišlja	1210. zaljubljen
260. iskustvom	577. ni jednu jedinu	894. razmišljaju	1211. zanima
261. ispoštuje	578. nijednu zamerku	895. raznovrsne	1212. zanimljive
262. ispravnih	579. nije dosadno	896. raznovrsno	1213. zanimljiva
263. istakao	580. nije isti	897. razuman	1214. zanimljivom
264. isti	581. nije loš	898. razumem	1215. zanimljivost
265. istrajnosti	582. nije loša	899. razumete	1216. zanimljivostima
266. istražuju	583. nije nadmen	900. razumeti	1217. žao
267. izađe u susret	584. nije naivan	901. razumevanju	1218. zapamti
268. izdavanja	585. nije najnekorektniji	902. razumnih	1219. zapamtiti
269. izdvojila	586. nije opsednuta	903. razvaljuje	1220. za primer
270. izgrađenim	587. nije preterano	904. razveseli	1221. za primer
271. izgradili	588. nije škrt	905. razvio	1222. zaronim
272. iznenađenje	589. nije škrt	906. razvoju	1223. zaslužno
273. izuzetna	590. nije strog	907. realan	1224. zaslužio
274. izuzetni	591. nije toliko loš	908. realnih	1225. zaslužio
275. izuzetnom	592. nije za baciti	909. red	1226. zaslužna
276. izvanredan	593. nikad dosadno	910. redak	1227. zaslužnih
277. izvanredna	594. nikakvih problema	911. redovan	1228. zauvek
278. izvanredni	595. ni malo loš	912. redovno	1229. zavide
279. izvuče	596. nimalo sujetna	913. relevantna	1230. zavidnom
280. japanac	597. nisam imao problem	914. relevantni	1231. zavolite
281. jasan	598. nisam nikada primetio	915. rešava	1232. završimo
282. jasna	599. nisam očekivao	916. rešavanja	1233. zdrav
283. jasne	600. nisam osećala	917. retki	1234. želela
284. jasni	601. nisam pogledao	918. retkim	1235. želi
285. jasnijim	602. nisam primetila	919. retko	1236. želja
286. jasnim	603. nisam uspela	920. retorikom	1237. željan
287. jasnu	604. ništa loše	921. rezultat	1238. želje

288. jedina	605. nisu dosadna	922. riznica	1239. želji
289. jedinstven	606. nisu loše	923. rođen	1240. željna
290. jedinstven	607. nisu nezanimljiva	924. rođena	1241. zen
291. jedinstvena	608. nisu obavezna	925. rodio za	1242. živ
292. jedinstvenim	609. normalnih	926. saborce	1243. žive
293. jednaka	610. nove	927. šalama	1244. živim
294. jednako	611. novi duh	928. šale	1245. zmaj
295. jednini	612. nudi	929. salonski	1246. značaj
296. jednom u životu	613. obavezno	930. samopoštovanja	1247. značaja
297. jednostavan	614. obeleži	931. samostalno	1248. značajna
298. jednostavna	615. obezbedi	932. samouverenu	1249. znače
299. jednostavnim	616. obiluje	933. šansu	1250. značilo
300. jedva čekam	617. obiman	934. saradnja	1251. znajem
301. jedva čekamo	618. objasniće	935. saraduje	1252. znalca
302. kako-tako	619. objasnila	936. sa razlogom	1253. znanjima
303. kamo sreće	620. objasnio	937. šarma	1254. znaš
304. kao kod kuće	621. objašnjena	938. šarmantan	1255. znati
305. kao on	622. objektivni	939. savesna	1256. zvezdicom
306. karakterno	623. objektivno	940. savladamo	1257. ^_^
307. kolega	624. obožavam	941. savladate	1258. :*
308. komičan	625. obrazovana	942. savladavanju	1259. +
309. kompetentan	626. obrazovanog	943. savremen	1260. :)
310. komunicira	627. obrazuje	944. savršen	1261. 10
311. komunikaciju	628. očekujemo	945. savršena	1262. 1
312. komunikativnost	629. odgovara	946. savršeno	1263. 10ka
313. koncentraciju	630. odličan	947. savršenstvo	1264. 10-ka
314. koncizan	631. odlične	948. sažet	1265. 5
315. konkretan	632. odlični	949. sažeta	1266. 5ica
316. konkretna	633. odlični	950. sažeto	
317. konkretno	634. odličnog	951. sažvaće	

## F.2. Jedinstveni oblici negativnih reči u skupu za obučavanje

1. abnormalni	382. malo znanja	763. netolerantan	1144. previsoki
2. agresivnijeg	383. malu	764. ne traži	1145. prevode
3. ali ne	384. mane	765. ne treba propuštati	1146. prevrće
4. aljkava	385. manjka	766. netrpeljivost	1147. prezahtevna
5. alkohola	386. masne viceve	767. neukusno	1148. prezauzet
6. alkoholičar	387. menja	768. neumesne	1149. prezentacija
7. alkoholisanom	388. menja	769. neupotrebljiva	1150. priča isto
8. anarhija	389. menjajte	770. neupotrebljivi	1151. prikrije
9. angažuje	390. menjate	771. ne urade	1152. prilagoditi
10. apelujem	391. mentalno stanje	772. neusaglašeni	1153. primedba
11. apsurdno	392. mešovito g	773. neusklađenih	1154. primitivna
12. arogancija	393. metalika	774. neusklađenost	1155. primorana
13. arogancije	394. miljenike	775. ne uslugu	1156. primorani
14. arogantan	395. minimalan	776. ne uspe	1157. prinuđeni
15. arogantna	396. minus	777. ne valja	1158. prinudno
16. arogantno	397. misterija	778. nevaspitan	1159. priprosta
17. arogantnog	398. mizoginiju	779. nevaspitana	1160. pritiska
18. bahata	399. mlaka	780. ne važi	1161. profesorčić
19. bahatost	400. mlati	781. nevažnim	1162. proizvoljno
20. baš briga	401. mnogo	782. ne vlada	1163. projektora
21. baš najzanimljivij	402. mnogo boljih	783. ne vole	1164. promašaj
22. beleške	403. moć	784. ne volim	1165. promašen
23. besa	404. modernizuje	785. ne volimo	1166. promašila
24. beskoristan	405. mogao bi	786. ne voliš	1167. promašio
25. besmisla	406. mogao bi	787. ne vredi	1168. promene raspoloženj
26. besvesti	407. mogla bi	788. ne za	1169. promeniti
27. bez	408. mogla bi	789. ne zadovoljava	1170. promenljive
28. beže	409. monolog	790. nezadovoljavajuće	1171. promenljivog
29. beži	410. monopol	791. nezadovoljna	1172. propustićeš
30. bez ikakve	411. monopola	792. nezainteresovan	1173. proračunljiv
31. bez imalo	412. monoton	793. nezainteresovana	1174. prosečnim
32. bežite	413. monotono	794. nezainteresovani	1175. prosek
33. bez jasnih	414. monotonost	795. nezaintersovan	1176. prošlo vreme
34. bez komentara	415. morala	796. nezanimljiv	1177. protivrečna
35. bez mozga	416. morao bi	797. nezanimljiva	1178. protivzakonit
36. bezobrazna	417. možda bi	798. nezanimljivi	1179. psovanje
37. bez odgo varajućil	418. može biti	799. nezanimljivo	1180. psuje
38. bez pojma	419. može to	800. ne zaslužuje	1181. psujete
39. bez posla	420. mrmolja	801. nezavisnu	1182. pudlica
40. bez poštovanja	421. mrtvačka	802. ne žele	1183. puke

41. bez razumevanja	422. mrzi	803. ne želi	1184. puno
42. bez reda	423. mržnju	804. ne želim	1185. puno bolje
43. bezuspešno	424. mučan	805. nezgodan	1186. puno raditi
44. bezveze	425. muči	806. nezgodna	1187. pun sebe
45. bez veze	426. muke	807. ne zna	1188. puž
46. bezvezne	427. muljažom	808. neznalica	1189. radije
47. bezvoljno	428. mutljavina	809. ne znam	1190. rakija
48. bezvredna	429. mutna	810. neznanja	1191. ranije
49. bi bilo poželjno	430. muževan	811. neznanju	1192. ranijih godina
50. bije	431. nabacan	812. ne žuri	1193. rano
51. bi mogao	432. nabacuje	813. nigde	1194. rasejana
52. bi moglo	433. nabeđen	814. nije ažurirano	1195. rasejani
53. bingo	434. nabeđenim	815. nije baš	1196. raspoloženja
54. bisera	435. nabiflaju	816. nije bitan	1197. rat
55. biserka	436. nabrajanjima	817. nije bitno	1198. ravna linija
56. bi trebala	437. nabuba	818. nije dao	1199. razglabao
57. bi trebao	438. nabusita	819. nijedno	1200. različita
58. bizaran	439. nadležne	820. nijednu	1201. različiti
59. blati	440. nadmen	821. nije dobar	1202. razočarali
60. blokada	441. nadmen	822. nije dolazila	1203. razočaran
61. bojažljiv	442. nadobudan	823. nije dosadno	1204. razočarao
62. bolela	443. nadobudna	824. nije dosledan	1205. razočarenje
63. bolesnih	444. nadrždana	825. nije dostojan	1206. realnije
64. bolesno	445. na engleskom	826. nije dostupan	1207. recepte
65. bolje da	446. nafrakana	827. nije dovoljno	1208. rekla-kazala
66. boluje	447. nagađamo	828. nije dozvoljeno	1209. reklama
67. bombardovao	448. na granici	829. nije držao	1210. remetio
68. božanstva	449. nahodenju	830. nije inspirativan	1211. reprodukuje
69. bradi	450. nahvatati	831. nije interesantan	1212. retardirani
70. briše	451. najdosadnije	832. nije ispričao	1213. retko drži
71. bruka	452. najdosadniji	833. nije istina	1214. retko kad
72. bruka	453. najdosadniji	834. nije izvežbala	1215. retko ko
73. brzinom munje	454. najgori	835. nije jača strana	1216. retko postoji
74. brzo	455. najlošiji	836. nije komunikativan	1217. revidira
75. brzopleto	456. najlošijih	837. nije kooperativan	1218. reviziju
76. bubače	457. najluđa	838. nije korektna	1219. rigidna
77. buke	458. najmanje	839. nije korektno	1220. robot
78. bulji	459. najmanju	840. nije kreativna	1221. roknete
79. butinu	460. najmiliji	841. nije loš	1222. rovari
80. ćefne	461. najmonotonije	842. nije među	1223. ruga
81. ćefu	462. najneomiljeniji	843. nije menjao	1224. rupe
82. čekala	463.	844. nije mogao	1225. ruši



	najneorganizovanija		
83. čekati	464. najomraženijih	845. nije mogla	1226. ružan
84. čemer	465. najteži	846. nije mu bitno	1227. ružnije
85. čemu svrha	466. naklonjeniji	847. nije najinteresantniji	1228. šablonu
86. cepidlači	467. nametnuti	848. nije napisana	1229. sahranu
87. cigaretu	468. namiguje	849. nije naučio	1230. sakrijete
88. cigaru	469. namiguje	850. nije neophodno	1231. sama sastavim
89. cima	470. namrgođen	851. nije normalna	1232. samo da
90. ciničan	471. namršteno	852. nije očigledno	1233. samokontrola
91. čita	472. namučio	853. nije odgovarao	1234. samoprozvani
92. čitanja	473. naopačke	854. nije od koristi	1235. samo sebe
93. čitao	474. napadna	855. nije odmakla	1236. samo toliko
94. copy/paste	475. napakovani	856. nije održao	1237. selektivno
95. crtanje	476. na papiru	857. nije omogućio	1238. senilan
96. čudi	477. napeta	858. nije ostvarila	1239. sepoku
97. čudni	478. napiše	859. nije otvoren	1240. siledžija
98. čudno	479. naplaćuje	860. nije pokazao	1241. široko
99. čudno	480. naporan	861. nije posvećena	1242. sirovina
100. čudnog	481. naporna	862. nije predavač	1243. sitničarenje
101. čudnom	482. naporno	863. nije predavao	1244. sitničavo
102. čuva	483. naprežete	864. nije predviđena	1245. sitnice
103. :d	484. napušta	865. nije pričala	1246. sitno
104. daleko	485. napustila	866. nije pripremila	1247. šizo frenije
105. daleko	486. napustite	867. nije redovno	1248. skidanje
106. daleko bilo	487. na stolu	868. nije rođena	1249. skine
107. daleko iznad	488. na strani	869. nije shvatila	1250. skinuto
108. dalje komentarisala	489. navijen	870. nije siguran	1251. skraćeno
109. danguba	490. nažalost	871. nije sigurna	1252. skripte
110. đavo	491. na zub	872. nije skrenuo	1253. skripti
111. dedovina	492. neadekvatan	873. nije spreman	1254. skrnavljenje
112. dekokracija	493. neadekvatan	874. nije spremna	1255. skupo
113. dekolteima	494. neadekvatna	875. nije stimulativan	1256. slab
114. delovao	495. ne aktivnosti	876. nije stručan	1257. slaba
115. depresivan	496. ne angažuje	877. nije teška	1258. slabašna
116. dere	497. neargumentovano	878. nije učio	1259. slabih
117. detaljnije	498. neažurna	879. nije ujednačeno	1260. slabo
118. digao	499. neažurnost	880. nije umeo	1261. slajdovima
119. diktiranje	500. ne bavi	881. nije usklađen	1262. slatkorečiva
120. dim	501. ne bih dao	882. nije u skladu	1263. slepi
121. disidenta	502. nebitna	883. nije u stanju	1264. slobode

122. diskriminiše	503. ne bi trebao	884. nije važan	1265. složeni
123. distanciran	504. nebulozno	885. nije vešt	1266. smanji
124. dlaku u jajetu	505. nebulozu	886. nije zanimljiv	1267. smanjite
125. dna	506. neće baviti	887. nije zanimljivo	1268. smarač
126. dno	507. neće držati	888. nije želeo	1269. smarao
127. dobacuje	508. neće imati	889. nikada	1270. smejanja
128. dokaže	509. neće izgraditi	890. nikako	1271. smejati
129. doper	510. neće moći	891. nikakva	1272. šmeke
130. dosadna	511. ne ceni	892. nikakva komunikacija	1273. smorena
131. dosadna	512. nećeš položiti	893. nikakvi	1274. snađem
132. dosadnom	513. neće trebati	894. nikakvog smisla	1275. snađi
133. dosađuje	514. nečovjek	895. nikakvu komunikaciju	1276. snaja
134. dospela	515. ne čuju	896. ni korisno	1277. snishodljiv
135. dostupniji	516. ne da	897. nimalo	1278. spava
136. dotična	517. ne da se	898. ni moguć	1279. spavaju
137. dotična	518. ne dobije	899. ni najmanje	1280. spopada
138. dotične	519. ne dolazi	900. ni najosnovnije	1281. sporija
139. dotični	520. nedopustivo	901. ni objaviti	1282. sporije
140. doviđenja	521. nedosledna	902. nipodaštavanje	1283. sprečava
141. dramom	522. nedostaje	903. ni približno	1284. sputava
142. dramu	523. nedostatak	904. nisam gotivila	1285. sram
143. drastično	524. nedostojan	905. nisam naučila	1286. sramota
144. drhtanju	525. nedostupna	906. nisam naučio	1287. sramotnim
145. drska	526. nedostupni	907. nisam ništa	1288. sramotno
146. drski	527. nedovoljno	908. nisam oduševljen	1289. sramotnu
147. drsko	528. ne dovršava	909. nisam stekla	1290. štetan
148. drskost	529. nedovršene	910. nisam ubeđen	1291. stidljiva
149. drugačija	530. ne dozvoljava	911. nisam umeo	1292. strah
150. drugačiji	531. ne dozvoljavajući	912. nisam uspeo	1293. strahodisciplini
151. drugih ocena	532. nedozvoljeno	913. nisam utvrdila	1294. stranih
152. drugu sliku	533. ne drže	914. nisam zadovoljna	1295. strašnije
153. drzak	534. ne drži	915. nisi dobar	1296. strašnom
154. dubok	535. ne edukovali	916. niske	1297. stražari
155. đubre	536. neenergičnosti	917. niske	1298. strepi
156. duva	537. neetični	918. ništa	1299. stres
157. dvosmislena	538. nefer	919. ništa bolji	1300. stresa
158. dvostrukim	539. nefleksibilan	920. ništa informativnija	1301. stresan
159. džaba	540. nefleksibilnost	921. ništa korisno	1302. stroga
160. ekscentričan	541. negativne	922. ništa lepo	1303. strogi

161. ekscentričnog	542. ne gleda	923. ništa naučiti	1304. strogo
162. eksplozivna	543. ne govori	924. ništa pametno	1305. stroža
163. esejskih	544. nehumano	925. ništa posebno	1306. stručnijeg
164. fale	545. ne ide	926. niste slušali	1307. stručnijoj
165. fantom	546. neinformisan	927. niste spremni	1308. stvor
166. figurira	547. neinspirativna	928. nisu adekvatne	1309. subjektivan
167. flertuje	548. neinteresantni	929. nisu baš	1310. sudim
168. folira	549. ne iskazuje	930. nisu bila	1311. šugava
169. folirant	550. ne ispituje	931. nisu interesantni	1312. sugestivna
170. folirantkinja	551. ne ispunjava	932. nisu interesovali	1313. sujeta
171. formalnostima	552. neizgrađen	933. nisu mogli	1314. sujetan
172. foruma	553. neživljen	934. nisu moguće	1315. sujetniju
173. forumu	554. ne izlazi	935. nisu monotona	1316. suknji
174. frapantno	555. nejasan	936. nisu najbolje	1317. suknjici
175. frlja	556. nejasna	937. nisu najsrećiji	1318. sulud
176. frustracije	557. nejasni	938. nisu obavezne	1319. sumnjivi
177. gafovima	558. nejasno	939. nisu objektivne	1320. šuplja
178. galame	559. nejasnoće	940. nisu obrađeni	1321. suprotne
179. glad za	560. nejednakost	941. nisu oslobađajući	1322. šurnaest
180. glasna	561. neki	942. nisu potrebne	1323. suviše
181. glasno	562. nekoegzistentan	943. nisu preterano	1324. suvoparan
182. glumac	563. nekompetentni	944. nisu realne	1325. suvoparno
183. glupe	564. nekomunikativan	945. nisu tako	1326. suvoparnost
184. glupim	565. nekomunikativna	946. nisu u skladu	1327. svašta
185. glupost	566. ne kontam	947. nisu u stanju	1328. svemoguća
186. glupostima	567. nekooperativna	948. nisu zadovoljni	1329. sve osim
187. gluvi	568. nekorektno	949. nisu zanimljiva	1330. svira
188. google	569. nekorektnog	950. nisu zaslužile	1331. svog sina
189. gora	570. nekorektnost	951. ni upisivala	1332. taj
190. gord	571. nekorektnost	952. ni zabavno	1333. tajnosti
191. greši	572. nekorisna	953. noćna mora	1334. taksista
192. greškama	573. nekorisni	954. noćni izlazak	1335. takvima
193. grozan	574. nekorisno	955. noge	1336. tamo neki
194. grozna	575. nekoristan	956. nonšalantan	1337. tatinog sina
195. grozno	576. ne koristi	957. nonšalantnost	1338. tek
196. gruba	577. nekreativna	958. novcem	1339. tek tolika
197. gura	578. nekritička	959. novina	1340. tek treba
198. gusto	579. nekvalitetna	960. nula	1341. telefonira
199. haos	580. ne liči	961. obarala	1342. teorijska
200. haotične	581. neljubazan	962. obarani	1343. tera
201. haotično	582. neljubazan	963. obaveza	1344. teško

202. histerična	583. neljubaznost	964. obavezna	1345. teškom
203. hitno	584. ne luta	965. obavezne	1346. težak
204. hladna	585. nema	966. obaveznih	1347. teže
205. horor	586. nema	967. obavezno	1348. tih
206. horor	587. nema argumenata	968. obesim	1349. tihi
207. hrvatskom	588. nema interakcije	969. obići	1350. tiho
208. hvala bogu	589. nema iskustva	970. obiman	1351. tik
209. hvali	590. nema izgrađen	971. obimni	1352. toliko i
210. identično	591. nema jasan	972. obmanom	1353. toliko o
211. idiot	592. nema jasne	973. obnavljao	1354. tone
212. ignorišete	593. nemaju neke	974. obogaljio	1355. tračarenja
213. igra	594. nemaju pravo	975. obrazloži	1356. tračevima
214. igraju	595. nemaju puno smisla	976. očaj	1357. tragično
215. igre	596. nemaju veze	977. očajan	1358. traju
216. ikada	597. nemaju želju	978. očajno	1359. trapava
217. improvizovao	598. nema komunikaciju	979. očekivao	1360. traumatično
218. improvizuje	599. nema koristi	980. odbija	1361. traže
219. inati	600. nema kvalitet	981. ode	1362. trčanje
220. insistira	601. nema materijala	982. odlazak	1363. treba
221. insistirala	602. nema materijala	983. odlazi	1364. treba da
222. interes	603. nema mesta	984. odlutam	1365. trebala bi
223. irelevantnu	604. nemam ništa	985. odmaže	1366. trebalo bi
224. iritantan	605. nema mnogo	986. odokativno	1367. trebaš
225. iritantna	606. nema mogućnosti	987. odokativnom	1368. trenutnim raspoloženjem
226. iritira	607. nemam reči	988. odrađuje	1369. troše
227. ironija	608. nema ni	989. odrekne	1370. tuga
228. iscedi	609. nema odgovarajuću	990. odskače	1371. tu i tamo
229. iščekivanje	610. nema određen	991. odstranite	1372. tukli
230. iscrpljen	611. nema opravdanja	992. odstupa	1373. tumor
231. iseckanih	612. nema pravo	993. odsustvo	1374. tumore
232. isfrustriran	613. nema prepisivanja	994. odsutna	1375. turistička
233. isfrustrirana	614. nema progledavanje	995. odustali	1376. tužno
234. iskompleksiran	615. nema rada	996. ogluvele	1377. ubačenih
235. iskompleksiran	616. nema sile	997. ogromna	1378. ubijem
236. iskompleksirana	617. nema strategiju	998. okreću	1379. ubili

237. ispade	618. nemate ni	999. okriivi	1380. ubilo
238. ispadi	619. nema više	1000. omalovažava	1381. ucnom
239. ispadima	620. nema viziju	1001. omalovažavanje	1382. udavi
240. ispaštaće	621. nema želju	1002. omanja	1383. ufitiljio
241. ispeče	622. ne mnogo	1003. onemoćuca	1384. ugleda
242. ispitaju	623. ne mnogo omiljen	1004. onemoćuca	1385. uhvatio
243. ispituje	624. nemoguće	1005. onemoćućen	1386. ujednačen
244. ispostavi	625. nemojte	1006. onemoćućeno	1387. uklopi
245. ispuca	626. ne mora	1007. onemoćućio	1388. ukrštenicu
246. išta	627. nemotivisana	1008. opasna	1389. ulizica
247. istiće	628. nemotivisanost	1009. opasno	1390. ulizuje
248. istu priču	629. ne može	1010. oprezu	1391. ulizuju
249. isuviše	630. ne možemo	1011. opširan	1392. umanjuju
250. izadem	631. ne možeš	1012. opušteniji	1393. umesto
251. izbačen	632. ne nauče	1013. oružje	1394. umislila
252. izbaci	633. ne nudi	1014. osnovac	1395. umišlja
253. izbacuje	634. neobaveštenost	1015. osnovne	1396. umišljen
254. izbegava	635. ne objasni	1016. osnovne škole	1397. umišljena
255. izbegavajte	636. ne objašnjava	1017. osnovnoj	1398. umoran
256. izbegavati	637. ne objavljuje	1018. osoran	1399. umre
257. izbegavati	638. neobjektivno	1019. osrednje	1400. unazadi
258. izdanak	639. ne obraća pažnju	1020. ostaje	1401. uništavaju
259. izdere	640. ne obrazlaže	1021. ostavlja	1402. u oblacima
260. izgovorom	641. neobrazovnu	1022. oštrije	1403. uobražen
261. izgube	642. ne očekujući	1023. osuđivanja	1404. uobražena
262. izgubi kontrolu	643. ne ocenjuje	1024. osvetoљubiv	1405. uplašio
263. izgubila	644. ne odgovara	1025. otaljavanje	1406. urlanje
264. izgubio	645. neodgovoran	1026. otera	1407. u širinu
265. izgubljena	646. ne odgovori	1027. otežava	1408. u slovo
266. izgubljena	647. neodgovorna	1028. otkači	1409. uspavaju
267. izigrava	648. neodgovorna	1029. otresito	1410. uspavan
268. izivljavanjem	649. ne održava	1030. otrov	1411. uspavanka
269. iz klupe izašli	650. ne održavaju	1031. otužna	1412. uspavaš
270. izlišna	651. neodštampanog	1032. ova	1413. uspavljujući
271. izmislio	652. ne odustaje	1033. ova	1414. uspavljujuća
272. iznad svih	653. neophodna	1034. ozbiljan	1415. usporen
273. iznervira	654. neoprostivo	1035. .p	1416. uštedite
274. iznervirana	655. ne ostaje	1036. paćenik	1417. uštogljen
275. izostaje	656. neozbiljnost	1037. padala	1418. ustoličen
276. izveštačen	657. ne pametna	1038. padao	1419. uvrede
277. izveštačena	658. nepismen	1039. padneš	1420. uvredila
278. izvini	659. nepismena	1040. pala	1421. užasan

279. izvlačenjem	660. nepodesnog	1041. pali	1422. užasna
280. izvlačili	661. nepodnošljivo	1042. palite	1423. užasno
281. izvrgava	662. ne pojavljuje	1043. papagaj	1424. uz svesku
282. jad	663. ne pokazuju	1044. paranoična	1425. valjalo bi
283. jadni	664. ne poklapaju	1045. pare	1426. varira
284. jedno	665. ne pokušava	1046. pazite	1427. važan
285. jedno	666. ne polazi za rukom	1047. pecanje	1428. veličina
286. jasnija	667. ne položite	1048. penzionisati	1429. više
287. jede	668. ne pomaže	1049. pijan	1430. vičući
288. jednoličan	669. ne pominjite	1050. pitanje je	1431. vikanje
289. jednolično	670. ne poslušajte	1051. plače	1432. vike
290. jednu te istu	671. nepoštovanje	1052. plaše	1433. više voli
291. još manje	672. ne poštuje	1053. plašljiv	1434. više založi
292. kada bi	673. nepotizmu	1054. plašljiva	1435. više znaju
293. kako-kad	674. nepotpuni	1055. poboljšati	1436. viši
294. kamoli	675. nepotrebna	1056. početnik	1437. visina
295. kamo sreće	676. nepotrebne	1057. podcenjuje	1438. viski
296. kancer	677. nepotrebno	1058. podela	1439. vraća
297. kapriciozan	678. ne povezuje	1059. podeljen	1440. vređanje
298. karleušu	679. ne poznaje	1060. podigao je nos	1441. vrti
299. kasni	680. ne poznaje	1061. podigne	1442. xx
300. kasni	681. nepoznavanje	1062. podložan	1443. za*****
301. kasno	682. nepraktična	1063. podmukao	1444. zabludu
302. katastrofa	683. nepravedan	1064. podsmehom	1445. zaboravim
303. katastrofalna	684. nepravičan	1065. pogrešio	1446. zaboravite
304. katastrofalno	685. neprecizna	1066. pogrešna	1447. zabranjena
305. kazni	686. nepredvidiva	1067. pogrešnom	1448. zabune
306. kinji	687. nepregledno	1068. pogube	1449. zadržt
307. kitnjast	688. ne prenosi	1069. pogubljen	1450. zadržtih
308. kliktanje	689. ne prenosi	1070. pojavio	1451. zadržte
309. koliko i	690. ne preporučujem	1071. pojede	1452. zaglupljujućih
310. koliko toliko	691. ne prepuštajte	1072. pokajao	1453. zagorčaju
311. komediju	692. ne preterano	1073. pokupe	1454. zagorčam
312. kompleksan	693. ne pridržava	1074. političar	1455. zahteva
313. kompleksi	694. neprijatan	1075. politički obojena	1456. zahtevaju
314. kompleksniji	695. neprijateljski	1076. politike	1457. zahtevnih
315. komplikovan	696. neprijatno	1077. pomeriti	1458. zahtevniji
316. komplikovanije	697. neprikladni	1078. pomirite sa činjenicom	1459. zakasni
317. konflikta	698. neprilagođena	1079. pomračina	1460. zakida
318. konflikte	699. ne priliči	1080. poništite	1461. žale
319. konfuzna	700. ne primenjuje	1081. ponižava	1462. žalite

320. konfuzne	701. neprimeren	1082. popravili	1463. žalosno
321. konfuzno	702. neprimerena	1083. poraditi	1464. zalutala
322. kontradiktorni	703. neprimerene	1084. poražavajuće	1465. zalutao
323. kontrolom	704. neprimereno	1085. pošalje	1466. zameniti
324. kopirane	705. nepripremljena	1086. posle	1467. zamerio
325. korisnije	706. ne pristaje	1087. poso	1468. zamerke
326. korumpirano	707. nepristojan	1088. poštapalice	1469. zanemarljiva
327. kosi	708. nepristupačan	1089. postavio	1470. zanese
328. košmar	709. ne privlači	1090. potcenjivački	1471. zanimljivija
329. kozo	710. neprofesionalac	1091. potcenjivanje	1472. zanimljiviji
330. kraj	711. neprofesionalna	1092. potencira	1473. žao
331. krajnost	712. neprofesionalni	1093. pothitnu	1474. zaostaju
332. kratka	713. neprofesionalno	1094. potpitanja	1475. zapali
333. kratki	714. neprofesionalnost	1095. potrebna	1476. zapetlja
334. kratkih živaca	715. ne pruži	1096. potruđi	1477. zapita
335. kriminalna	716. ne radimo	1097. povlači	1478. zapostavlja
336. kritika	717. ne raspože	1098. povrh svega	1479. zaseban
337. kritikovati	718. neraspoloženo	1099. površni	1480. zaspao
338. kritiku	719. nerazgovetno	1100. povuče	1481. zastarelih
339. krivac	720. nerazuman	1101. pozer	1482. zastarelim
340. krut	721. ne razumeju	1102. pozornici	1483. zastraši
341. kruta	722. ne razumem	1103. prase	1484. zastrašujuć
342. kukajte	723. nerazumevanje	1104. pravi	1485. zategnut
343. kupiti	724. nerazumno	1105. pravi važan	1486. zatupi
344. kupuju	725. nerealan	1106. prebacih	1487. zatvoren
345. kurs	726. nerealna	1107. prebacili	1488. zatvorenicima
346. kvar	727. nerealne	1108. prebrzo	1489. zauzela
347. kvazi	728. nerealno	1109. predrasuda	1490. zauzet
348. laž	729. nervira	1110. predugo	1491. završe
349. lažira	730. nervni slom	1111. preispitana	1492. zaziru
350. lažni	731. nervozan	1112. prekine	1493. zbijenog
351. leči	732. nervoze	1113. preklapanje	1494. zbrda-zdola
352. lenje	733. nervozno	1114. preko bundeva	1495. zbrza
353. lenji	734. ne šaljite	1115. preko noći	1496. zbunjujući
354. leti	735. nesamostalna	1116. prekršena	1497. zevanje
355. liče	736. nesavesna	1117. prekuca vaju	1498. zezneš
356. lično	737. ne sećam	1118. preleće	1499. zeznuto
357. ljigavo	738. ne shvatajući	1119. preotela	1500. zgadi
358. loš	739. ne shvatate	1120. prepisao	1501. zgaditi
359. loše	740. nesiguran	1121. prepisuje	1502. zgroženo
360. loše	741. nesigurna	1122. prepotentan	1503. živahan
361. lošem	742. ne simpatiše	1123. prepotentna	1504. zla

362. loši	743. nesistematična	1124. prepotentni	1505. zlo
363. loših	744. ne sluša	1125. prepušteni	1506. zločin
364. lošija	745. ne slušam	1126. presedan	1507. zloupotreba
365. lošije	746. ne sme	1127. presmešan	1508. zlovoljni
366. lucifer	747. ne smeje	1128. prestala	1509. žurio
367. ludila	748. ne smemo	1129. prestrog	1510. zvaničan
368. ludnicu	749. nesnađenim	1130. pretera	1511. zveraju
369. lupa	750. nesporazuma	1131. preterane	1512. zvezdaš
370. lupi	751. nespremna	1132. preterivanje	1513. zzzz
371. lutrija	752. nespremnost	1133. preteruje	1514. :(
372. magarci	753. ne stigne	1134. pretežak	1515. :(
373. malaksalo	754. nestrpljiva	1135. pretrčavala	1516. -1
374. male	755. nestručan	1136. preuzetih	1517. -10
375. maligni	756. nestručna	1137. prevara	1518. -100
376. malo	757. nesvesno	1138. prevarant	1519. 2
377. malo bolja	758. ne sviđa	1139. prevaziđen	1520. 200 na sat
378. malo bolje	759. ne sviđi	1140. prevaziđeno	1521. 80-tih
379. malo češće	760. nesvrishodnosti	1141. prevelike	
380. malo korektniji	761. netačne	1142. preveliki	
381. malo više	762. netaľentovana	1143. previše	



## F.3. Jedinstveni oblici pozitivnih sentiment reči u skupu za testiranje

1. adekvatnu	205. komunikativan	409. obraća	613. šansu
2. aktivno	206. konkretna	410. obrađeno	614. saradivati
3. aktuelno	207. konkretna	411. obratiti	615. saraduje
4. aktuelnostima	208. konkretno	412. obrazovan	616. šareni
5. altruista	209. kontinuitet	413. obrazovana	617. saslušati
6. angažuje	210. korektan	414. odgovara	618. savesno
7. autoritet	211. korektni	415. odgovarajuće	619. savet
8. ažuriranu	212. korektnija	416. odgovoran	620. saveta
9. ažurna	213. korektnije	417. odgovoriće	621. savete
10. ažurno	214. korektnijeg	418. odgovorila	622. savladali
11. bez problema	215. korektno	419. odlične	623. savremena
12. bez smaranja	216. korektnom	420. odlični	624. savremenim
13. bez zamerke	217. korektnost	421. odličnih	625. savršen
14. bitnijim	218. koriguje	422. odlično	626. sažvakano
15. bitno	219. korisne	423. odličnom	627. sećam
16. blago	220. korisnim	424. odmerenošću	628. sećate
17. bolja	221. korisno	425. održi	629. sećati
18. boljeg	222. korisnu	426. oduševilo	630. shvatio
19. bolju	223. kralj	427. oduševio	631. shvatljiv
20. car	224. kraljica	428. oduševljeni	632. sigurno
21. car	225. kraljina	429. odvoja	633. simpatičan
22. carica	226. krasni	430. ogromno	634. simpatični
23. carina	227. kratkim	431. ok	635. širokih
24. čast	228. kratko	432. olakšavaju	636. širokog
25. čeka	229. kulturnan	433. omiljen	637. širokog spectra
26. cenim	230. kulture	434. omiljeni	638. sistematičan
27. cenjen	231. kulturna	435. omogućava	639. sistematično
28. čini	232. kulturnog	436. opuštena	640. sistematizaciju
29. čista	233. kulturom	437. opuštenijih	641. sjajna
30. čista	234. kupio	438. organizovana	642. sjajnim
31. citirati	235. kvaliteta	439. organizuju	643. skidam kapu
32. čitko	236. kvalitetan	440. osloboditi	644. skroman
33. čudo	237. kvalitetnih	441. oslovljava	645. skypa
34. čuje	238. kvalitetnija	442. osmehom	646. slikoviti
35. čvrsto	239. kvalitetno	443. ostvorena	647. slikovito
36. id	240. kvalitetu	444. ostvarljivo	648. sloboda
37. daje	241. laf	445. osvestiti	649. slušajući
38. da je više	242. lako	446. osveženje	650. smernice
39. daleko	243. legenda	447. osvrtom	651. snađu
40. da nije	244. legenda	448. otvorena	652. snova

41. da nije nje	245. legende	449. otvoreno	653. solidan
42. dekana	246. lep	450. otvorenost	654. solidna
43. deset	247. lepoj	451. ozbiljan	655. sposobni
44. desetka	248. lepom	452. pamćenje	656. spremni
45. desetkom	249. lik	453. pametan	657. srce
46. detaljne	250. ljubav	454. pametan	658. srdačna
47. detaljnih	251. ljubavi	455. pametna	659. sreća
48. detaljno	252. ljubaznim	456. pamte	660. sreće
49. direktna	253. ljudina	457. pamtite	661. srećni
50. divan	254. ljudina	458. pamtiti	662. srećno
51. divim	255. ljudski	459. pažljiv	663. sreću
52. dobri	256. majka	460. pedagog	664. staložen
53. dobrih	257. majstor	461. pedagoškim	665. stara škola
54. dobro	258. maksimalno	462. perfekcionista	666. strog
55. dobrom	259. maksimalnom	463. perfektan	667. strpljiv
56. dobronamerna	260. malo je takvih	464. perspektivna	668. strpljiva
57. dodatne	261. malo ovakvih	465. pevajući	669. strpljivo
58. do detalja	262. maštoviti	466. plodom	670. stručna
59. dogovor	263. mesto više	467. plus	671. stručnjak
60. dogovoru	264. metodičan	468. podeli	672. stručno
61. dogurati	265. mila	469. podiglo	673. stub
62. donese	266. mir	470. podržala	674. super
63. doprinosi	267. miran	471. podsticanje	675. super
64. dorasla	268. mlada	472. podstiče	676. susretljiv
65. dosledna	269. mlađeg	473. pohale	677. suština
66. dosta	270. mnogo	474. pohvale	678. suštini
67. dostupnost	271. mnogo	475. pohvalno	679. suštinsko
68. dovodi	272. mnoštvom	476. pojednostavi	680. sve od sebe
69. dovoljno	273. moderna	477. pokrije	681. svestran
70. drag	274. mogućnost	478. pokušava	682. svestran
71. dragulja	275. mogućnosti	479. polako	683. svestrana
72. dr house	276. moralno	480. pomagala	684. svestrani
73. drug	277. motivacione	481. pomažu	685. svetla
74. drugom	278. motivator	482. pomoći	686. svetla
75. drži	279. motivisana	483. pomogao	687. svetla tačka
76. dugogodišnjim	280. mozak	484. pomogla	688. svetskih
77. duhovit	281. može	485. pomogle	689. svetu
78. duhovita	282. načitana	486. ponosi	690. svidaju
79. duhoviti	283. načitanosti	487. ponosna	691. tačno
80. duhovitosti	284. nadaren	488. ponovi	692. talentovana
81. edukacije	285. naglasak	489. ponovljeni	693. temeljna
82. efikasna	286. naglasi	490. popravljaju	694. tišinu
83. efikasnijim	287. najbitnija	491. posavetuje	695. tolerantni

84. efikasno	288. najbolje	492. posebno	696. traži
85. ekstra	289. najbolje	493. posećenosti	697. trenutno
86. ekstra	290. najboljeg	494. posluži	698. truda
87. enciklopedija	291. najboljem	495. postane	699. trude
88. energična	292. najbolju	496. pošten	700. trudio
89. energija	293. najdivniji	497. poštovanje	701. trudom
90. energijom	294. najinovativniji	498. poštovanjem	702. trudu
91. entuzijastično	295. najinspirativnija	499. poštuje	703. tutor
92. entuzijastu	296. najinteligentnija	500. posvećena	704. tutorijale
93. entuzijizam	297. najinteresantija	501. posvećeni	705. ubedljiv
94. entuzijazma	298. najjači	502. posvećeno	706. ubedljivo
95. entuzijazmom	299. najkorektnija	503. posvećenog	707. uče
96. fantastičan	300. najkorektnija	504. posvećuje	708. učestvuje
97. fantastičan	301. najkorektniji	505. posvećuje	709. uči
98. fantastična	302. najkorektniji	506. posveti	710. učio
99. fantastično	303. najkorisnije	507. posvetio	711. učiti
100. fascinantna	304. najkorisnijih	508. potrebne	712. udeli
101. fenomenalan	305. najkulturnija	509. potruditi	713. udomio
102. fenomenalan	306. najkvalitetniji	510. poučna	714. udžbenika
103. fenomenalna	307. najlepša	511. poučnim	715. ugled
104. fer	308. najlepše	512. poveže	716. uključenost
105. fina	309. najosnovnije	513. povezuje	717. ulaže
106. fina	310. najpametnija	514. pozitivne	718. u malom prstu
107. fino	311. najpametnijih	515. pozitivnu	719. ume
108. fleksibilan	312. najperspektivniji	516. poznaje	720. unapređenju
109. fluentnim	313. najposvećeniji	517. poznavanju	721. upečatljiva
110. forsira	314. najpozitivnijih	518. praksu	722. urade
111. frajerčina	315. najpoznatiji	519. praktična	723. uredno
112. garantujem	316. najpristupačniji	520. praktične	724. urodi
113. garda	317. najpristupačnijih	521. praktični	725. u roku
114. general	318. najrealniji	522. praktično	726. usklađivala
115. genijalna	319. najširim	523. praktičnu	727. usmeravanje
116. genije	320. najsystematičniji	524. prava	728. uspe
117. gospodin	321. najslađa	525. pravi	729. uspešne
118. gostujuće	322. najsporija	526. pravi	730. uspeva
119. gromadu	323. najstručniji	527. pravih	731. u srž
120. grupno	324. najstručnijih	528. pravi način	732. usvojeno
121. harizma	325. najtemeljniji	529. pravio	733. uticao
122. harizmatična	326. najvažnije	530. precizan	734. uvažava
123. harizmatičnost	327. najvažnijih	531. precizno	735. uvek
124. harizmu	328. najveće	532. predan	736. uvek je tu
125. hiper- korektna	329. najveći	533. predivna	737. uživa

126. hoće	330. najveći	534. predivno	738. uživala
127. humor	331. najzabavniji	535. predusretljiv	739. uzvatiće
128. hvala	332. najzanimljivija	536. predusretljiva	740. važni
129. hvala	333. najzanimljivija	537. pregledno	741. važnih
130. idealan	334. najzanimljiviji	538. prenošenju	742. veću
131. ideje	335. najzanimljivijih	539. prepametan	743. vedar
132. idol	336. naklon	540. prepuna	744. vedrog
133. idol	337. na konju	541. preuzeti	745. veličina
134. ima	338. napor	542. prezadovoljan	746. veliki
135. imati još	339. napravi	543. prezadovoljna	747. veliki
136. individualnu	340. napred	544. pričljiva	748. velikom
137. informisani	341. naša	545. prijatan	749. vesela
138. informiše	342. nasmejali	546. prijatelj	750. vešt
139. insistirao	343. nasmeje	547. prijateljski	751. vešto
140. inspirativan	344. nastavi	548. prijatna	752. video
141. inspirativan	345. nastoji	549. prijatni	753. videti
142. inspirativna	346. nauče	550. prijatniji	754. više takvih
143. inspirativnih	347. naučićete	551. prilagođava	755. visokom
144. institucija	348. naučni	552. prilagođen	756. visoku
145. integritetom	349. ne bi izgubio	553. priliku	757. visprena
146. inteligencije	350. ne boji	554. primeniti	758. vlada
147. inteligentan	351. neće iznervirati	555. primenljivo	759. volela
148. interakcija	352. ne dozvoljava	556. primer	760. voljan
149. interaktivnim	353. nedvosmisleno	557. primere	761. vredni
150. interesantan	354. nejcenjenijih	558. primerima	762. vrednovanje
151. interesantne	355. ne maltretira	559. pripremljen	763. vrh
152. interesantni	356. nema mnogo milost	560. prisniji	764. vrhunskom
153. interesantni	357. nemam problem	561. pristojan	765. vrsan
154. interesantnijim	358. nemam zamerke	562. pristojna	766. vrsni
155. interesovanje	359. nema prepisivanja	563. pristupačniji	767. vrstan
156. interesovanjima	360. ne možete zameriti	564. prisutan	768. zabavan
157. interesuju	361. neograničenog	565. privilegijom	769. zabavno
158. intresantna	362. neopisiv	566. privilegovano	770. zadovoljan
159. iskrenu	363. neponovljiva	567. privilegovanon	771. zadovoljna
160. iskustava	364. neprocenjivog	568. privlačnijim	772. zagarantovano
161. iskustvom	365. ne promakne	569. prizemljena	773. zahvaljujući
162. ispoštujem	366. nesebično	570. prođeš	774. zahvaljujući
163. ispredaje	367. nesumnjivo	571. profesionalan	775. zahvalnosti
164. isprobati	368. nesvakidašnji	572. profesionalna	776. zainteresovani
165. istim aršinom	369. ne teoretiše	573. profesionalni	777. zainteresovanost
166. istinu	370. ne toleriše	574. prolazite	778. zalaganje
167. izaberem	371. neverovatan	575. promocije	779. zalaže
168. izdigne	372. neverovatnu	576. promovisanju	780. zaljubljenik

169. izdvaja	373. neviđeni	577. propisano	781. zanima
170. iznad	374. ni do ramena	578. prosečan	782. zanimljive
171. iznadprosečnih	375. nije bila takva	579. proširuje	783. zanimljivije
172. iznese	376. nije desilo	580. prosto	784. zanimljiviji
173. izuzetan	377. nije dosadno	581. prostora	785. zanimljivim
174. izuzetni	378. nije kriva	582. pruži	786. zanimljivom
175. izuzetno	379. nije loša	583. prvima	787. zanimljivost
176. izvanredan	380. nije merilo	584. prvom	788. zanimljivu
177. izvanredan	381. nije naporan	585. psiholog	789. zapamćeno
178. izvanredna	382. nije odbijao	586. punim pravom	790. zapamti
179. izvanredno	383. nije odbila	587. rade	791. zapamtiti
180. izvanrednog	384. nije slepo držao	588. radi	792. zaposlio
181. izvrsna	385. nije strašan	589. radionice	793. za poželeti
182. izvrsnu	386. nije sujetan	590. raspolaganju	794. zasluženno
183. izvrstan	387. nije suvoparno	591. rastura	795. zaslužila
184. jak	388. nije tako loš	592. razgovorom	796. zauvek
185. jaka	389. nije teška	593. različite	797. zavidim
186. jasan	390. nije teško	594. razmišljaju	798. zavidnom
187. jasne	391. nije toliko loša	595. razmišljamo	799. završio
188. jasni	392. nije u srodstvu	596. raznim	800. zdravo
189. jasnije	393. nikad na užrb	597. razumem	801. žele
190. jasnim	394. nisam propustila	598. razumevanje	802. želim
191. javlja	395. nismo pogubili	599. razumevanju	803. željom
192. jedina	396. nisu copy-paste	600. razumljiv	804. zlatna
193. jedina	397. nisu striktno	601. razumljive	805. značajno
194. jedinstvena	398. normale	602. razumna	806. znači
195. jedna je	399. novu	603. razumno	807. znala
196. jednostavni	400. obaveštavaju	604. razvijaju	808. znalac
197. jesu	401. obimno	605. redak	809. znanjem
198. još bolji	402. objasnio	606. redovno	810. znaš
199. kad god	403. objašnjavajući	607. rector	811. znate
200. kampovala	404. objašnjenja	608. relevantno	812. zove
201. klasa	405. objašnjenje	609. rešenje	813. 10
202. ključno	406. objektivan	610. retka	814. 100
203. kompetentna	407. objektivno	611. retoričar	815. 5ica
204. kompetentna	408. obožavamo	612. samostalno	816. 8

## F.4. Jedinствени oblici negativnih sentiment reči u skupu za testiranje

1. agresivno	232. loši	463. neslaganja	694. prepisivanje
2. apsurdno	233. loših	464. ne slažem	695. prepisuje
3. arogancije	234. lošija	465. ne sluša	696. prepotentna
4. arogantan	235. lošim	466. ne sme	697. prepustio
5. arogantan	236. lucidan	467. nesposobna	698. prestorga
6. arogantna	237. ma	468. nespretnih	699. pretenciozan
7. arogantne	238. mala	469. nestabilan	700. preterivanje
8. aršine	239. malo	470. nestabilna	701. pretnju
9. autoritativan	240. malo znanja	471. ne stigne	702. pretvaraju
10. baca	241. maltretiraju	472. nestručan	703. prevaziđenog
11. bacite	242. mane	473. nestručan	704. prevazišle
12. bahato	243. manje	474. nestručna	705. previše
13. bahatost	244. manje	475. netačne	706. previše
14. bahatosti	245. manje od	476. ne tiču	707. prevoda
15. barijere	246. manje znaju	477. ne trudi	708. prevrtljiva
16. bauka	247. marksizam	478. ne uče	709. prezahtevan
17. beskorisan	248. maše	479. ne uči	710. prezaštićen
18. beskorisne	249. maskota	480. neukih	711. prezire
19. beskorisnih	250. mator	481. ne ulaže	712. prgav
20. beskorisnim	251. minus	482. ne ume	713. prikrije
21. beskorisno	252. mlada	483. ne umem	714. prikriva
22. beskoristan	253. mnogo	484. neumesan	715. prinuđen
23. beskrajno	254. mnogo	485. neumesnom	716. privatne
24. besmislena	255. mobilnim	486. ne unapređuje	717. privatnim
25. besmislene	256. mogao bi	487. neupotrebljiva	718. privrženi
26. besmisleno	257. mogao bi	488. ne upoznaješ	719. problem
27. besna	258. momente	489. ne uspeva	720. problemi
28. bespotrebno	259. monolozima	490. ne usvaja	721. proces
29. bez	260. monoton	491. ne uvažava	722. prodaje
30. bezdušna	261. monotono	492. nevaljalac	723. proganja
31. bez komentara	262. mora da	493. nevaspitana	724. promašila
32. bezobrazan	263. morala bi	494. neverovatno	725. promenljiv
33. bezobrazno	264. moraš	495. ne vole	726. prosečnih
34. bezosećajan	265. može da	496. ne voli	727. proširiti
35. bezrazložno	266. mrgud	497. ne vredi	728. prošlosti
36. bez znanja	267. mrmlja	498. nezadovoljna	729. prostih
37. bilo ko	268. mrtvu	499. nezainteresovana	730. protekcije
38. bi trebao	269. mrtvu volje	500. nezainteresovana	731. psuje
39. blati	270. mrze	501. nezainteresovano	732. puši
40. bolja	271. mučenje	502.	733. pustio korene

41. brza	272. muči	503. ne zanimaju	734. random
42. brže	273. muti	504. nezanimljive	735. rasejan
43. brzo	274. nabacan	505. ne zaslužuje	736. rasejani
44. brzom	275. nabuba	506. ne želim	737. različite
45. bubaju	276. nadmen	507. nezgodna	738. razlikuju
46. bubanjem	277. nadmena	508. ne zna	739. rođak
47. budale	278. nadmeno	509. ni do kolena	740. rodbinska
48. buka	279. nadobudnosti	510. ni interesantna	741. rugalice
49. burno	280. nadržanog	511. nije	742. ruši
50. cepa	281. najbolji	512. nije baš	743. ružna
51. cepidlači	282. najdosadnija	513. nije bilo prilike	744. ružnih
52. cigaretu	283. najgluplji	514. nije bio	745. sadista
53. cinik	284. najgore	515. nije bio	746. sama
		zadovoljan	
54. čitač	285. najgoreg	516. nije bitno	747. samoinicijativno
55. čitanje	286. najgori	517. nijednom	748. samoljubac
56. copy/paste	287. najgorih	518. nijednom pojavio	749. samopromo više
57. čudan	288. najgorim	519. nije dobra	750. samo ukoliko
58. čudi	289. najlošiji	520. nije dobro	751. samovoljno
59. čuditi	290. najlošijih	521. nije dosadno	752. širili
60. čudne	291. najmanje	522. nije dostupan	753. siromaštvo
61. čuvajte	292. najnekorektnija	523. nije dovoljno	754. sitničav
62. da bar	293. najnižom	524. nije dovoljno	755. skače
63. da bi	294. nakačila	525. nije iskusan	756. sklepana
64. daleko bilo	295. namazana	526. nije korektan	757. sklonite
65. daleko od	296. namerno	527. nije korektno	758. skreće
66. davi	297. nameti	528. nije korisna	759. skripti
67. definitivno nije	298. na momente	529. nije mogao	760. slab
68. dekolte	299. napad	530. nije moguće	761. slabija
69. dete	300. napada	531. nije najbolji	762. slabim
70. devijantno	301. naporni	532. nije nas naučio	763. slabo
71. diktirao	302. na račun	533. nije neko	764. slajdove
72. diskutabilni	303. nasrće	534. nije normalno	765. smanjio
73. dno	304. nasumice	535. nije objektivan	766. smanjivati
74. dno	305. nasumično	536. nije pregledan	767. smanjuje
75. dopuniti	306. naterajte	537. nije raspoložena	768. smarača
76. dosadi	307. navodno	538. nije realan	769. smešan
77. dosadna	308. nažalost	539. nije siguran	770. smešno
78. dosadna	309. nažvrlja	540. nije spremna	771. snishodljivo
79. dosadne	310. ne	541. nije stigao	772. sociopati
80. dosadnih	311. neadekvatan	542. nije tačno	773. specifičan
81. dosadnim	312. neartikulisano	543. nije upućena	774. spora

82. dosadno	313. ne baš	544. nije u stanju	775. sporije
83. dosta	314. nebeski	545. nije uvek	776. sprečen
84. doterane	315. ne bi	546. nije voljan	777. spusti na zemlju
85. dotična	316. ne bih poželela	547. nije za	778. spustiti
86. drama	317. ne bi mogao	548. nije zanimljivo	779. sramotno
87. drčna	318. ne bi razumeo	549. nikad	780. srednjih
88. drkoš	319. ne bi smelo	550. nikada	781. srednjoj
89. drski	320. nebitne	551. nikad niko	782. srednju
90. dvojka	321. nebitnom	552. nikako	783. srušio
91. dvostruke	322. nebuloza	553. nikakva	784. stara
92. dvostruke standarde	323. neće	554. nikakve	785. starački
93. ekscentričan	324. neće biti	555. nikakve veze	786. stare
94. falio	325. neće da	556. ni korisna	787. starog
95. favorizuje	326. neće održati	557. ni kučeta	788. staromodno
96. filozofske	327. nećete dobro	558. ni mačeta	789. šta znam
97. formalista	328. ne čita	559. nimalo simpatičan	790. stegnut
98. forsiranja	329. neću	560. ni naročito	791. stoka
99. frustrirani	330. neću dozvoliti	561. nipodaštavanja	792. strašno
100. glumata	331. ne čuje	562. nipodaštavanjem	793. strašno
101. glumi	332. ne daje	563. ni predavanja	794. strave
102. glup	333. ne davi	564. ni približno	795. štreber
103. glupe	334. ne dešava	565. ni razumevanja	796. striktan
104. glupostima	335. ne dođe	566. nisam čula	797. striktna
105. gora	336. ne dolaze	567. nisam imao zamerki	798. strog
106. gost	337. ne dolaziti	568. nisam video	799. strog
107. greši	338. ne dopada	569. nisam više	800. stroga
108. greška	339. ne doprinosi	570. niskom	801. strogi
109. greškama	340. nedopustivo	571. nisku	802. strogoćom
110. greški	341. nedostaje	572. nismo ni	803. strogo g
111. haos	342. nedostatak	573. nismo videli	804. strožeg
112. haotičan	343. nedostataka	574. ništa lepo	805. subjektivno
113. hendikepirala	344. nedostatke	575. ništa novo	806. subjektivnog
114. hira	345. nedostatku	576. ništa organizovanc	807. subjektivnom
115. histeriše	346. nedostojan	577. ništa pametno	808. sujeta
116. hitno	347. nedostojno	578. ništa pozitivno	809. sujetna
117. hladna	348. nedostupan	579. niste dobili	810. suknja
118. hm	349. nedostupni	580. nisu baš	811. sulude
119. hostesa	350. nedovoljna	581. nisu bitne	812. sumanute
120. hranljivo	351. ne dozvoljava	582. nisu način	813. suprotan
121. hrvatskih	352. negativne	583. nisu ni	814. suviše
122. hukće	353. ne gleda	584. nisu ništa	815. suvoparan



123. huškač	354. ne govori	585. nisu potrebna	816. suvoparni
124. hvalisavac	355. ne ići	586. nisu toliko	817. suvoparno
125. identične	356. ne ide u prilog	587. niti jasna	818. suza
126. ignoriše	357. ne idite	588. niti jednu	819. svađa
127. igrama na sreću	358. ne idu	589. niti zanimljiva	820. svako može
128. ikad	359. neinspirativna	590. niti zna	821. svega
129. ikada	360. neinteligentna	591. ni trunku	822. tajni agent
130. ikakvog	361. neinteresantan	592. ni u kakvoj	823. tak va kak va je
131. imbecili	362. neinteresantna	593. niža vrsta	824. talasnim dužinama
132. inačenje	363. ne interesuje	594. niže	825. tankog
133. insistiranju	364. neiskusna	595. nižu rasu	826. teatralnosti
134. irelevantne	365. ne izlaziti	596. nula	827. tek kada
135. iritantan	366. nejasna	597. obara	828. teletabisi
136. iritantan	367. nejasno	598. obaveza	829. teorijama
137. ironijom	368. ne javlja	599. obavezno	830. tera
138. iscedi	369. nekakvog	600. obesmislio	831. terao
139. isčitavanju	370. nekompetentni	601. obim	832. teret
140. isečci	371. nekompetentnih	602. obimna	833. teška
141. isfrustrirane	372. nekonkretno	603. obimnije	834. teško
142. isfrustriranom	373. nekorektan	604. obojena	835. težak
143. isfrustriranost	374. nekorektna	605. oborila	836. težak
144. iskaljuje	375. ne korektnim	606. oborio	837. težinu
145. isključivo	376. nekorektno	607. očajna	838. težu
146. iskompleksirana	377. nekorektnosti	608. očekuje	839. ti
147. iskorištavanje	378. nekorisna	609. odbio	840. tiho
148. ismejani	379. nekreativnom	610. ode	841. toliko o
149. ispaštali	380. ne kupovati	611. odgovorila	842. totalni
150. ispodprosečan	381. nelogično	612. odlazi	843. tragedija
151. ispodprosečnih	382. nema	613. odmogne	844. trapavih
152. isprepisivala	383. nema daljih	614. od reči do reči	845. traume
153. ista je	384. nema dovoljno	615. odu	846. treba da
154. istera	385. nema gori	616. ogovara	847. trenutnom raspoloženju
155. isto tako	386. nema jasan	617. ograničava	848. tupav
156. i tako	387. nema jasno	618. ograničava	849. tvrdo glava
157. iza leđa	388. nemaju prolaz	619. ograničeno	850. uf
158. izašao	389. nemam lepo	620. ogromne	851. ukinuo
159. izbacuje	390. nemam reči	621. okreće očima	852. umesto
160. izbegava	391. nemam šta	622. okrenu	853. umišljena
161. izbegavate	392. nema nadmen	623. okrenut	854. umišljena
162. izbegne	393. nema pojma	624. okvirni	855. uopšte
163. izbjije	394. nema pojma	625. olako	856. upire
164. izdiže	395. nema potrebe	626. omiljenu	857. upropaštava

165. izgovori	396. nemarnost	627. omražen	858. urlanje
166. izgubljen	397. nema šta	628. opada	859. u sebe
167. izgubljena	398. nemaštovito	629. opasna	860. usko grlo
168. izgubljeno	399. nema uspešnog	630. opravdanja	861. uskraćuju
169. izivljava	400. nemilosrdan	631. opterećuje	862. uslovljava
170. izivljavanje	401. ne mogu	632. ordinalna	863. utroši
171. izivljavao	402. nemoгуćnost	633. osiromašena	864. uvlaкуša
172. izlazi	403. ne mora	634. osnovcima	865. uvlaкуše
173. izlišnog	404. ne može	635. osrednjost	866. uvrede
174. izmešanom	405. ne možeš	636. oštro	867. užasa
175. izmišlja	406. ne možete	637. osuđeni	868. užasna
176. izmišljen	407. ne obavezuje	638. otaljava	869. užasni
177. iznad	408. neobičan	639. otežan	870. užasno
178. iznadprosečni	409. ne obilazi	640. otežavajuću	871. užasno
179. izvan	410. ne obraća	641. otvoreno pokazuje	872. uzvišena
180. izvredala	411. neodgovarajući	642. ova	873. varira
181. jadni	412. neodgovornost	643. ovaj	874. viceve
182. jaz	413. ne olakšava	644. ovog	875. videh
183. jedini	414. ne oprašta	645. pada	876. virne
184. jedinica	415. neorganizovana	646. pada u vodu	877. više fokusira
185. jedinicama	416. neorganizovanosti	647. padne	878. više uraditi
186. jedinice	417. nepedagog	648. pakosnik	879. visokim
187. jedino	418. nepodnošljiva	649. penzionisati	880. voli mlađe
188. jedno isto	419. ne podržava	650. persona	881. vraća
189. jednoličnim	420. ne pojasni	651. pesimizam	882. vreda
190. jednom po	421. ne pojavljuje	652. plaćaju	883. vrve
191. jednosmerna	422. ne položi	653. pobegao	884. zaboravio
192. jedva	423. ne pomažu	654. podeli	885. zabranio
193. još malo	424. ne ponaša	655. podseti	886. zabraniti
194. još manje	425. ne poseduje	656. pogrešne	887. zabunu
195. kako dune	426. nepostojanje	657. pojavi	888. za čega
196. kardinalne	427. ne postoje	658. pojavila	889. zahtevi
197. kasne	428. nepoštovanje	659. pojavio	890. zahtevna
198. kasni	429. ne poštuju	660. pokajala	891. zahtevna
199. katastrofa	430. ne posvećivanje	661. pokvaren	892. zakida
200. katastrofa	431. ne posvećuje	662. ponavlja	893. žale
201. katastrofalno	432. nepotrebim	663. ponavljao	894. žalim
202. kazni	433. nepoveravati	664. ponižavaju	895. žaliti
203. kerber	434. nepoznavanje	665. ponižavanje	896. žalost
204. klimav	435. ne preduzima	666. ponižavanje	897. zaluta
205. klonite	436. nepredvidiv	667. popiše	898. zalutao
206. kompleks	437. ne pregleda	668. popravimo	899. zamara

207. konfuzija	438. nepregledne	669. poradi	900. zamera
208. konfuzne	439. ne pričam	670. poraditi	901. zameram
209. kontroverzna	440. neprijateljsku	671. porede	902. zamerka
210. kopirani	441. neprikladna	672. pošalje	903. zaostavština
211. kopiranog	442. neprikladne	673. potceni	904. zapita
212. korigovati	443. nepripremljena	674. potcenjivačke	905. zapitaš
213. korumpiranim	444. nepristojna	675. potcenjuje	906. zasmetalo
214. košta	445. nepristupačnosti	676. površna	907. zastarela
215. koštao	446. neprofesionalan	677. povukao	908. zauzet
216. krajnost	447. neprofesionalno	678. požali	909. zauzima
217. kraju	448. ne pušta	679. prebegli	910. završava
218. kratak fitilj	449. nerad	680. preceni	911. zazirala
219. kratka	450. ne radi	681. precrtate	912. zbrkano
220. kritički	451. ne razume	682. predosadni	913. zbunjen
221. kritikuje	452. ne razumem	683. predškolskog	914. zbunjena
222. krivice	453. ne razumem	684. predugo	915. zdravstveno stanje
223. krivo	454. ne razumete	685. preferira	916. zelen
224. kuburi	455. nerazumljivi	686. preispitati	917. zgurano
225. lascivne	456. nerazumljivo	687. prekida	918. živaca
226. lažima	457. nerazumno	688. preko veze	919. zlo
227. lečenje	458. nervoza	689. prenaduvano	920. zloupotrebe
228. ljubomornih	459. ne saznaje	690. prenatrpana	921. žuri
229. loša	460. nesigurna	691. preobiman	922. 0
230. loša	461. nesigurnost	692. preobimno	923. 1969
231. lošem	462. ne širi	693. preokrenuo	

## **G. Intenzifikatori sentimenta identifikovani u korpusu**

Tokom anotacije, anotatori su izdvajali reči koje jačaju ili slabe iskazani sentiment. Ove reči su navedene u nastavku.

1. malo	25. uvek	49. pravi	73. prepuna
2. zaista	26. ubedljivo	50. prava	74. prepuno
3. nešto	27. apsolutno	51. čisto	75. prepune
4. blago	28. preterano	52. čista	76. maksimalno
5. pomalo	29. ogromna	53. često	77. neizmerno
6. u suštini	30. ogroman	54. sve u svemu	78. najviše
7. totalno	31. dosta	55. relativno	79. premalo
8. ekstremno	32. stvarno	56. samo	80. nemerivog
9. baš	33. više	57. ponekad	81. nemerivo
10. prilično	34. stalno	58. u principu	82. visokog
11. krajnje	35. neverovatno	59. ikad	83. visok
12. krajnja	36. neverovatna	60. ikada	84. visoki
13. još kako	37. neverovatan	61. ikakvog	85. prototip
14. veoma	38. konstantno	62. određeni	86. donekle
15. previše	39. konstantan	63. mnogo	87. daleko
16. iznimno	40. definitivno	64. jako	88. oličenje
17. potpuno	41. jako puno	65. izuzetno	89. sušto
18. potpuna	42. jako veliki	66. vanserijski	90. itekako
19. jako	43. neviđeni	67. jako	91. suviše
20. više nego	44. neviđena	68. isključivo	92. Enormno
21. pun	45. neviđeno	69. najveći	93. potpuno
22. puno	46. svakako	70. potpuna	94. potpuni
23. puna	47. mnogo toga	71. vrlo	95. uopšte
24. retko	48. skroz	72. sasvim	

## **H. Domensko specifične fraze identifikovane u korpusu**

Tokom anotacije anotatori su izdvajali specifične fraze koje studenti učestalo koriste u recenziranju nastavnog osoblja. U nastavku je priložena lista domensko specifičnih fraza.

- |                             |                               |
|-----------------------------|-------------------------------|
| 1. jedan/na od retkih       | 22. tačan kao sat             |
| 2. jedan/na od najboljih    | 23. u svakom pogledu          |
| 3. jedan/na od najgorih     | 24. svaka mu čast             |
| 4. jedan/na od boljih       | 25. svaka joj čast            |
| 5. jedan/na od malo         | 26. svaka čast                |
| 6. šta reći, a ne zaplakati | 27. pregazilo vreme           |
| 7. gde bi nam bio kraj      | 28. jednom rečju              |
| 8. gde bi nam kraj bio      | 29. sve od sebe               |
| 9. gleda kroz prste         | 30. sve pohvale               |
| 10. progledala kroz prste   | 31. starog kova               |
| 11. bez komentara           | 32. porcija je jednaka za sve |
| 12. izašli iz klupe         | 33. mlati prazna slama        |
| 13. sve samo ne             | 34. jedva čekam               |
| 14. može to bolje           | 35. bez sumnje                |
| 15. može to mnogo bolje     | 36. skidam kapu               |
| 16. od reči do reči         | 37. dlaku u jajetu            |
| 17. stara škola             | 38. samo reči hvale           |
| 18. svih vremena            | 39. zreo za (penziju npr)     |
| 19. na svetu                | 40. ne polazi za rukom        |
| 20. na celom fakultetu      | 41. brzinom munje             |
| 21. kapa dole               |                               |

## **I. Ključne reči negacije identifikovane u korpusu**

Tokom anotacije anotatori su označavali ključne reči negacije koje negiraju određeni deo sadržaja. U nastavku je priložena lista jedinstvenih oblika negacije.

**Tabela L1 Jedinstveni oblici ključnih reči negacije i frekventnost njihovog pojavljivanja u korpusu**

<b>Oblik negacije</b>	<b>Broj pojavljivanja</b>
ne	793
nije	290
nema	124
nisu	85
ni	80
nisam	56
neće	27
nit i	24
nismo	16
nemaju	14
nemam	13
niste	11
nećete	9
ništa	8
neću	7
niko	6
nemojte	4
nit	3
nemate	3
nećemo	3
nijednu	3
nikada	2
nikad	2
nisi	2
nećeš	2
nijedan	1
nikakvu	1
nikako	1
nemamo	1
nemaš	1
nikakav	1

## J. Modeli klasifikacije sentimenta zasnovani na algoritmima nadgledanog učenja

U analizi sentimenta primenom algoritama mašinskog učenja razvijeni su modeli klasifikacije primenom *Naïve Bayes*, *Support vector machines* i *k-Nearest Neighbor* algoritama. U nastavku je prikazana optimalna konfiguracija svakog algoritma, koja daje najbolje rezultate. Modeli su razvijeni u RapidMiner alatu.

### J.1. Naïve Bayes klasifikator

```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<process version="5.3.015">
  <context>
    <input/>
    <output/>
    <macros/>
  </context>
  <operator activated="true" class="process" compatibility="5.3.015" expanded="true" name="Process">
    <process expanded="true">
      <operator activated="false" class="retrieve" compatibility="5.3.015" expanded="true" height="60"
        name="Retrieve wordlist" width="90" x="112" y="390">
        <parameter key="repository_entry" value="wordlist"/>
      </operator>
      <operator activated="false" class="text:process_document_from_file" compatibility="5.3.002" expanded="true" height="76"
        name="Process Documents from Files (2)" width="90" x="246" y="390">
        <list key="text_directories">
          <parameter key="pozitivno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\TEST_SKUP\POZ"/>
          <parameter key="negativno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\TEST_SKUP\NEG"/>
        </list>
        <parameter key="file_pattern" value="*.*/>
        <parameter key="extract_text_only" value="false"/>
        <parameter key="encoding" value="UTF-8"/>
        <parameter key="keep_text" value="true"/>
        <parameter key="prune_method" value="absolute"/>
        <parameter key="prune_below_percent" value="5.0"/>
        <parameter key="prune_above_percent" value="100.0"/>
        <parameter key="prune_below_absolute" value="2"/>
        <parameter key="prune_above_absolute" value="9999"/>
      </operator>
      <operator activated="true" class="text:transform_cases" compatibility="5.3.002" expanded="true" height="60"
        name="Transform Cases (2)" width="90" x="45" y="30"/>
      <operator activated="true" class="text:tokenize" compatibility="5.3.002" expanded="true" height="60"
        name="Tokenize (2)" width="90" x="180" y="30"/>
      <operator activated="true" class="text:filter_stopwords_dictionary" compatibility="5.3.002" expanded="true" height="76"
        name="Filter Stopwords (2)" width="90" x="313" y="30">
        <parameter key="file" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\stopReci_Hrvatski.txt"/>
        <parameter key="encoding" value="UTF-8"/>
      </operator>
      <connect from_port="document" to_op="Transform Cases (2)" to_port="document"/>
      <connect from_op="Transform Cases (2)" from_port="document" to_op="Tokenize (2)" to_port="document"/>
      <connect from_op="Tokenize (2)" from_port="document" to_op="Filter Stopwords (2)" to_port="document"/>
      <connect from_op="Filter Stopwords (2)" from_port="document" to_port="document 1"/>
      <portSpacing port="source_document" spacing="0"/>
      <portSpacing port="sink_document 1" spacing="0"/>
      <portSpacing port="sink_document 2" spacing="0"/>
    </process>
  </operator>
  <operator activated="false" class="retrieve" compatibility="5.3.015" expanded="true" height="60" name="Retrieve model"
    width="90" x="447" y="345">
    <parameter key="repository_entry" value="model"/>
  </operator>
  <operator activated="false" class="apply_model" compatibility="5.3.015" expanded="true" height="76" name="Apply Model (2)"
    width="90" x="581" y="390">
    <list key="application_parameters"/>
  </operator>
  <operator activated="true" class="text:process_document_from_file" compatibility="5.3.002" expanded="true" height="76"
    name="Process Documents from Files" width="90" x="45" y="120">
    <list key="text_directories">
      <parameter key="pozitivno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\POZ"/>
      <parameter key="negativno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\NEG"/>
    </list>
  </operator>

```

```

<parameter key="file_pattern" value="*.*"/>
<parameter key="extract_text_only" value="false"/>
<parameter key="encoding" value="UTF-8"/>
<parameter key="keep_text" value="true"/>
<parameter key="prune_method" value="absolute"/>
<parameter key="prune_above_percent" value="100.0"/>
<parameter key="prune_below_absolute" value="2"/>
<parameter key="prune_above_absolute" value="9999"/>
<process expanded="true">
  <operator activated="true" class="text:transform_cases" compatibility="5.3.002" expanded="true" height="60"
    name="Transform Cases" width="90" x="45" y="30"/>
  <operator activated="true" class="text:tokenize" compatibility="5.3.002" expanded="true" height="60" name="Tokenize"
    width="90" x="180" y="30"/>
  <operator activated="true" class="text:filter_stopwords_dictionary" compatibility="5.3.002" expanded="true" height="76"
    name="Filter Stopwords (Dictionary)" width="90" x="315" y="30">
    <parameter key="file" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\stopReci_Hrvatski.txt"/>
    <parameter key="encoding" value="UTF-8"/>
  </operator>
  <operator activated="true" class="text:generate_n_grams_terms" compatibility="5.3.002" expanded="true" height="60"
    name="Generate n-Grams (Terms)" width="90" x="447" y="30"/>
  <operator activated="true" class="text:filter_by_length" compatibility="5.3.002" expanded="true" height="60"
    name="Filter Tokens (by Length)" width="90" x="581" y="30">
    <parameter key="min_chars" value="3"/>
  </operator>
  <connect from_port="document" to_op="Transform Cases" to_port="document"/>
  <connect from_op="Transform Cases" from_port="document" to_op="Tokenize" to_port="document"/>
  <connect from_op="Tokenize" from_port="document" to_op="Filter Stopwords (Dictionary)" to_port="document"/>
  <connect from_op="Filter Stopwords (Dictionary)" from_port="document" to_op="Generate n-Grams (Terms)" to_port="document"/>
  <connect from_op="Generate n-Grams (Terms)" from_port="document" to_op="Filter Tokens (by Length)" to_port="document"/>
  <connect from_op="Filter Tokens (by Length)" from_port="document" to_port="document 1"/>
  <portSpacing port="source_document" spacing="0"/>
  <portSpacing port="sink_document 1" spacing="0"/>
  <portSpacing port="sink_document 2" spacing="0"/>
</process>
</operator>
<operator activated="true" class="store" compatibility="5.3.015" expanded="true" height="60" name="Store wordlist"
  width="90" x="179" y="210">
  <parameter key="repository_entry" value="wordlist"/>
</operator>
<operator activated="true" class="x_validation" compatibility="5.3.015" expanded="true" height="112" name="Validation"
  width="90" x="313" y="120">
<process expanded="true">
  <operator activated="true" class="naive_bayes" compatibility="5.3.015" expanded="true" height="76" name="Naive Bayes"
    width="90" x="45" y="30"/>
  <connect from_port="training" to_op="Naive Bayes" to_port="training set"/>
  <connect from_op="Naive Bayes" from_port="model" to_port="model"/>
  <portSpacing port="source_training" spacing="0"/>
  <portSpacing port="sink_model" spacing="0"/>
  <portSpacing port="sink_through 1" spacing="0"/>
</process>
</operator>
<process expanded="true">
  <operator activated="true" class="apply_model" compatibility="5.3.015" expanded="true" height="76" name="Apply Model"
    width="90" x="45" y="30">
    <list key="application_parameters"/>
  </operator>
  <operator activated="true" class="performance_binominal_classification" compatibility="5.3.015" expanded="true" height="76"
    name="Performance" width="90" x="112" y="165">
    <parameter key="precision" value="true"/>
    <parameter key="recall" value="true"/>
    <parameter key="f_measure" value="true"/>
  </operator>
  <connect from_port="model" to_op="Apply Model" to_port="model"/>
  <connect from_port="test set" to_op="Apply Model" to_port="unlabelled data"/>
  <connect from_op="Apply Model" from_port="labelled data" to_op="Performance" to_port="labelled data"/>
  <connect from_op="Performance" from_port="performance" to_port="averagable 1"/>
  <portSpacing port="source_model" spacing="0"/>
  <portSpacing port="source_test set" spacing="0"/>
  <portSpacing port="source_through 1" spacing="0"/>
  <portSpacing port="sink_averagable 1" spacing="0"/>
  <portSpacing port="sink_averagable 2" spacing="0"/>
</process>

```



```
<connect from_op="Retrieve wordlist" from_port="output" to_op="Process Documents from Files (2)" to_port="word list"/>
<connect from_op="Process Documents from Files (2)" from_port="example set" to_op="Apply Model (2)" to_port="unlabelled data"/>
<connect from_op="Retrieve model" from_port="output" to_op="Apply Model (2)" to_port="model"/>
<connect from_op="Process Documents from Files" from_port="example set" to_op="Validation" to_port="training"/>
<connect from_op="Process Documents from Files" from_port="word list" to_op="Store wordlist" to_port="input"/>
<connect from_op="Validation" from_port="model" to_op="Store model" to_port="input"/>
<connect from_op="Validation" from_port="averagable 1" to_port="result 1"/>
<portSpacing port="source_input 1" spacing="0"/>
<portSpacing port="sink_result 1" spacing="0"/>
<portSpacing port="sink_result 2" spacing="0"/>
</process>
</operator>
</process>
```

## J.2. Support vector machines klasifikator

```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<process version="5.3.015">
  <context>
    <input/>
    <output/>
    <macros/>
  </context>
  <operator activated="true" class="process" compatibility="5.3.015" expanded="true" name="Process">
    <process expanded="true">
      <operator activated="false" class="retrieve" compatibility="5.3.015" expanded="true" height="60" name="Retrieve wordlist"
        width="90" x="112" y="390">
        <parameter key="repository_entry" value="wordlist"/>
      </operator>
      <operator activated="false" class="text:process_document_from_file" compatibility="5.3.002" expanded="true" height="76"
        name="Process Documents from Files (2)" width="90" x="246" y="390">
        <list key="text_directories">
          <parameter key="pozitivno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\TEST_SKUP\POZ"/>
          <parameter key="negativno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\TEST_SKUP\NEG"/>
        </list>
        <parameter key="file_pattern" value="*.*"/>
        <parameter key="extract_text_only" value="false"/>
        <parameter key="encoding" value="UTF-8"/>
        <parameter key="keep_text" value="true"/>
        <parameter key="prune_method" value="absolute"/>
        <parameter key="prune_below_percent" value="5.0"/>
        <parameter key="prune_above_percent" value="100.0"/>
        <parameter key="prune_below_absolute" value="2"/>
        <parameter key="prune_above_absolute" value="9999"/>
        <process expanded="true">
          <operator activated="true" class="text:transform_cases" compatibility="5.3.002" expanded="true" height="60"
            name="Transform Cases (2)" width="90" x="45" y="30"/>
          <operator activated="true" class="text:tokenize" compatibility="5.3.002" expanded="true" height="60" name="Tokenize (2)"
            width="90" x="180" y="30"/>
          <operator activated="true" class="text:filter_stopwords_dictionary" compatibility="5.3.002" expanded="true" height="76"
            name="Filter Stopwords (2)" width="90" x="313" y="30">
            <parameter key="file" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\stopReci_Hrvatski.txt"/>
            <parameter key="encoding" value="UTF-8"/>
          </operator>
          <connect from_port="document" to_op="Transform Cases (2)" to_port="document"/>
          <connect from_op="Transform Cases (2)" from_port="document" to_op="Tokenize (2)" to_port="document"/>
          <connect from_op="Tokenize (2)" from_port="document" to_op="Filter Stopwords (2)" to_port="document"/>
          <connect from_op="Filter Stopwords (2)" from_port="document" to_port="document 1"/>
          <portSpacing port="source_document" spacing="0"/>
          <portSpacing port="sink_document 1" spacing="0"/>
          <portSpacing port="sink_document 2" spacing="0"/>
        </process>
      </operator>
      <operator activated="false" class="retrieve" compatibility="5.3.015" expanded="true" height="60" name="Retrieve model"
        width="90" x="447" y="345">
        <parameter key="repository_entry" value="model"/>
      </operator>
      <operator activated="false" class="apply_model" compatibility="5.3.015" expanded="true" height="76" name="Apply Model (2)"
        width="90" x="581" y="390">
        <list key="application_parameters"/>
      </operator>
      <operator activated="true" class="text:process_document_from_file" compatibility="5.3.002" expanded="true" height="76"
        name="Process Documents from Files" width="90" x="45" y="120">
        <list key="text_directories">
          <parameter key="pozitivno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\POZ"/>
          <parameter key="negativno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\NEG"/>
        </list>
        <parameter key="file_pattern" value="*.*"/>
        <parameter key="extract_text_only" value="false"/>
        <parameter key="encoding" value="UTF-8"/>
        <parameter key="keep_text" value="true"/>
        <parameter key="prune_method" value="absolute"/>
        <parameter key="prune_above_percent" value="100.0"/>
        <parameter key="prune_below_absolute" value="2"/>
        <parameter key="prune_above_absolute" value="9999"/>
        <process expanded="true">
          <operator activated="true" class="text:transform_cases" compatibility="5.3.002" expanded="true" height="60" name="Transform Cases"
            width="90" x="45" y="30"/>
          <operator activated="true" class="text:tokenize" compatibility="5.3.002" expanded="true" height="60" name="Tokenize"
            width="90" x="180" y="30"/>
          <operator activated="true" class="text:filter_stopwords_dictionary" compatibility="5.3.002" expanded="true" height="76"
            name="Filter Stopwords (Dictionary)" width="90" x="315" y="30">
            <parameter key="file" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\stopReci_Hrvatski.txt"/>
            <parameter key="encoding" value="UTF-8"/>
          </operator>
        </process>
      </operator>
    </process>
  </operator>

```

```

<operator activated="true" class="text:generate_n_grams_terms" compatibility="5.3.002" expanded="true" height="60"
  name="Generate n-Grams (Terms)" width="90" x="447" y="30">
  <parameter key="max_length" value="1"/>
</operator>
<operator activated="true" class="text:filter_by_length" compatibility="5.3.002" expanded="true" height="60"
  name="Filter Tokens (by Length)" width="90" x="581" y="30">
  <parameter key="min_chars" value="3"/>
</operator>
<connect from_port="document" to_op="Transform Cases" to_port="document"/>
<connect from_op="Transform Cases" from_port="document" to_op="Tokenize" to_port="document"/>
<connect from_op="Tokenize" from_port="document" to_op="Filter Stopwords (Dictionary)" to_port="document"/>
<connect from_op="Filter Stopwords (Dictionary)" from_port="document" to_op="Generate n-Grams (Terms)" to_port="document"/>
<connect from_op="Generate n-Grams (Terms)" from_port="document" to_op="Filter Tokens (by Length)" to_port="document"/>
<connect from_op="Filter Tokens (by Length)" from_port="document" to_port="document 1"/>

  <portSpacing port="source_document" spacing="0"/>
  <portSpacing port="sink_document 1" spacing="0"/>
  <portSpacing port="sink_document 2" spacing="0"/>
</process>
</operator>
<operator activated="true" class="store" compatibility="5.3.015" expanded="true" height="60" name="Store wordlist"
  width="90" x="179" y="210">
  <parameter key="repository_entry" value="wordlist"/>
</operator>
<operator activated="true" class="x_validation" compatibility="5.3.015" expanded="true" height="112" name="Validation"
  width="90" x="313" y="120">
  <process expanded="true">
    <operator activated="true" class="support_vector_machine" compatibility="5.3.015" expanded="true" height="112"
      name="SVM" width="90" x="45" y="30"/>
    <connect from_port="training" to_op="SVM" to_port="training set"/>
    <connect from_op="SVM" from_port="model" to_port="model"/>
    <portSpacing port="source_training" spacing="0"/>
    <portSpacing port="sink_model" spacing="0"/>
    <portSpacing port="sink_through 1" spacing="0"/>
  </process>
  <process expanded="true">
    <operator activated="true" class="apply_model" compatibility="5.3.015" expanded="true" height="76" name="Apply Model"
      width="90" x="45" y="30">
      <list key="application_parameters"/>
    </operator>
    <operator activated="true" class="performance_binominal_classification" compatibility="5.3.015" expanded="true" height="76"
      name="Performance" width="90" x="112" y="165">
      <parameter key="precision" value="true"/>
      <parameter key="recall" value="true"/>
      <parameter key="f_measure" value="true"/>
    </operator>

    <connect from_port="model" to_op="Apply Model" to_port="model"/>
    <connect from_port="test set" to_op="Apply Model" to_port="unlabelled data"/>
    <connect from_op="Apply Model" from_port="labelled data" to_op="Performance" to_port="labelled data"/>
    <connect from_op="Performance" from_port="performance" to_port="averagable 1"/>
    <portSpacing port="source_model" spacing="0"/>
    <portSpacing port="source_test set" spacing="0"/>
    <portSpacing port="source_through 1" spacing="0"/>
    <portSpacing port="sink_averagable 1" spacing="0"/>
    <portSpacing port="sink_averagable 2" spacing="0"/>
  </process>
</operator>

  <operator activated="true" class="store" compatibility="5.3.015" expanded="true" height="60" name="Store model"
    width="90" x="447" y="75">
    <parameter key="repository_entry" value="model"/>
  </operator>
  <connect from_op="Retrieve wordlist" from_port="output" to_op="Process Documents from Files (2)" to_port="word list"/>
  <connect from_op="Process Documents from Files (2)" from_port="example set" to_op="Apply Model (2)" to_port="unlabelled data"/>
  <connect from_op="Retrieve model" from_port="output" to_op="Apply Model (2)" to_port="model"/>
  <connect from_op="Process Documents from Files" from_port="example set" to_op="Validation" to_port="training"/>
  <connect from_op="Process Documents from Files" from_port="word list" to_op="Store wordlist" to_port="input"/>
  <connect from_op="Validation" from_port="model" to_op="Store model" to_port="input"/>
  <connect from_op="Validation" from_port="averagable 1" to_port="result 1"/>
  <portSpacing port="source_input 1" spacing="0"/>
  <portSpacing port="sink_result 1" spacing="0"/>
  <portSpacing port="sink_result 2" spacing="0"/>
</process>
</operator>
</process>

```

### J.3. k-Nearest Neighbor klasifikator

```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<process version="5.3.015">
  <context>
    <input/>
    <output/>
    <macros/>
  </context>
  <operator activated="true" class="process" compatibility="5.3.015" expanded="true" name="Process">
    <process expanded="true">
      <operator activated="false" class="retrieve" compatibility="5.3.015" expanded="true" height="60"
        name="Retrieve wordlist" width="90" x="112" y="390">
        <parameter key="repository_entry" value="wordlist"/>
      </operator>
      <operator activated="false" class="text:process_document_from_file" compatibility="5.3.002" expanded="true" height="76"
        name="Process Documents from Files (2)" width="90" x="246" y="390">
        <list key="text_directories">
          <parameter key="pozitivno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\TEST_SKUP\POZ"/>
          <parameter key="negativno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\TEST_SKUP\NEG"/>
        </list>
        <parameter key="file_pattern" value="*.*"/>
        <parameter key="extract_text_only" value="false"/>
        <parameter key="encoding" value="UTF-8"/>
        <parameter key="keep_text" value="true"/>
        <parameter key="prune_method" value="absolute"/>
        <parameter key="prune_below_percent" value="5.0"/>
        <parameter key="prune_above_percent" value="100.0"/>
        <parameter key="prune_below_absolute" value="2"/>
        <parameter key="prune_above_absolute" value="9999"/>
        <process expanded="true">
          <operator activated="true" class="text:transform_cases" compatibility="5.3.002" expanded="true" height="60"
            name="Transform Cases (2)" width="90" x="45" y="30"/>
          <operator activated="true" class="text:tokenize" compatibility="5.3.002" expanded="true" height="60"
            name="Tokenize (2)" width="90" x="180" y="30"/>
          <operator activated="true" class="text:filter_stopwords_dictionary" compatibility="5.3.002" expanded="true"
            height="76" name="Filter Stopwords (2)" width="90" x="313" y="30">
            <parameter key="file" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\stopReci_Hrvatski.txt"/>
            <parameter key="encoding" value="UTF-8"/>
          </operator>
          <connect from_port="document" to_op="Transform Cases (2)" to_port="document"/>
          <connect from_op="Transform Cases (2)" from_port="document" to_op="Tokenize (2)" to_port="document"/>
          <connect from_op="Tokenize (2)" from_port="document" to_op="Filter Stopwords (2)" to_port="document"/>
          <connect from_op="Filter Stopwords (2)" from_port="document" to_port="document 1"/>
          <portSpacing port="source_document" spacing="0"/>
          <portSpacing port="sink_document 1" spacing="0"/>
          <portSpacing port="sink_document 2" spacing="0"/>
        </process>
      </operator>
      <operator activated="false" class="retrieve" compatibility="5.3.015" expanded="true" height="60" name="Retrieve model"
        width="90" x="447" y="345">
        <parameter key="repository_entry" value="model"/>
      </operator>
      <operator activated="false" class="apply_model" compatibility="5.3.015" expanded="true" height="76" name="Apply Model (2)"
        width="90" x="581" y="390">
        <list key="application_parameters">
        </operator>
      <operator activated="true" class="text:process_document_from_file" compatibility="5.3.002" expanded="true" height="76"
        name="Process Documents from Files" width="90" x="45" y="120">
        <list key="text_directories">
          <parameter key="pozitivno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\POZ"/>
          <parameter key="negativno" value="C:\Users\Korisnik\Dropbox\DOKTORAT\modeli\RapidMiner optimalni algoritam ML\komentar poz neg\NEG"/>
        </list>
        <parameter key="file_pattern" value="*.*"/>
        <parameter key="extract_text_only" value="false"/>
        <parameter key="encoding" value="UTF-8"/>
        <parameter key="keep_text" value="true"/>
        <parameter key="prune_method" value="absolute"/>
        <parameter key="prune_above_percent" value="100.0"/>
    </process>
  </operator>

```

```

<parameter key="prune_below_absolute" value="2"/>
<parameter key="prune_above_absolute" value="9999"/>
<process expanded="true">
  <operator activated="true" class="text:transform_cases" compatibility="5.3.002" expanded="true" height="60"
    name="Transform Cases" width="90" x="45" y="30"/>
  <operator activated="true" class="text:tokenize" compatibility="5.3.002" expanded="true" height="60"
    name="Tokenize" width="90" x="180" y="30"/>
  <operator activated="true" class="text:filter_stopwords_dictionary" compatibility="5.3.002" expanded="true" height="76"
    name="Filter Stopwords (Dictionary)" width="90" x="315" y="30">
    <parameter key="file" value="C:\Users\Rorisanik\Dropbox\DOKTORAT\model1\stopReci_Brvatski.txt"/>
    <parameter key="encoding" value="UTF-8"/>
  </operator>
  <operator activated="true" class="text:generate_n_grams_terms" compatibility="5.3.002" expanded="true" height="60"
    name="Generate n-Grams (Terms)" width="90" x="447" y="30"/>
  <operator activated="true" class="text:filter_by_length" compatibility="5.3.002" expanded="true" height="60"
    name="Filter Tokens (by Length)" width="90" x="581" y="30">
    <parameter key="min_chars" value="3"/>
  </operator>
  <connect from_port="document" to_op="Transform Cases" to_port="document"/>
  <connect from_op="Transform Cases" from_port="document" to_op="Tokenize" to_port="document"/>
  <connect from_op="Tokenize" from_port="document" to_op="Filter Stopwords (Dictionary)" to_port="document"/>
  <connect from_op="Filter Stopwords (Dictionary)" from_port="document" to_op="Generate n-Grams (Terms)" to_port="document"/>
  <connect from_op="Generate n-Grams (Terms)" from_port="document" to_op="Filter Tokens (by Length)" to_port="document"/>
  <connect from_op="Filter Tokens (by Length)" from_port="document" to_port="document 1"/>
  <portSpacing port="source_document" spacing="0"/>
  <portSpacing port="sink_document 1" spacing="0"/>
  <portSpacing port="sink_document 2" spacing="0"/>
</process>
</operator>
<operator activated="true" class="store" compatibility="5.3.015" expanded="true" height="60" name="Store wordlist"
  width="90" x="179" y="210">
  <parameter key="repository_entry" value="wordlist"/>
</operator>
<operator activated="true" class="x_validation" compatibility="5.3.015" expanded="true" height="112" name="Validation"
  width="90" x="313" y="120">
  <process expanded="true">
    <operator activated="true" class="k_nn" compatibility="5.3.015" expanded="true" height="76" name="k-NN"
      width="90" x="45" y="30">
      <parameter key="k" value="3"/>
      <parameter key="measure_types" value="NumericalMeasures"/>
      <parameter key="numerical_measure" value="CosineSimilarity"/>
    </operator>
    <connect from_port="training" to_op="k-NN" to_port="training set"/>
    <connect from_op="k-NN" from_port="model" to_port="model"/>
    <portSpacing port="source_training" spacing="0"/>
    <portSpacing port="sink_model" spacing="0"/>
    <portSpacing port="sink_through 1" spacing="0"/>
  </process>
  <process expanded="true">
    <operator activated="true" class="apply_model" compatibility="5.3.015" expanded="true" height="76" name="Apply Model"
      width="90" x="45" y="30">
      <list key="application_parameters"/>
    </operator>
    <operator activated="true" class="performance_binominal_classification" compatibility="5.3.015" expanded="true" height="76"
      name="Performance" width="90" x="112" y="165">
      <parameter key="precision" value="true"/>
      <parameter key="recall" value="true"/>
      <parameter key="f_measure" value="true"/>
    </operator>
    <connect from_op="Retrieve wordlist" from_port="output" to_op="Process Documents from Files (2)" to_port="word list"/>
    <connect from_op="Process Documents from Files (2)" from_port="example set" to_op="Apply Model (2)" to_port="unlabelled data"/>
    <connect from_op="Retrieve model" from_port="output" to_op="Apply Model (2)" to_port="model"/>
    <connect from_op="Process Documents from Files" from_port="example set" to_op="Validation" to_port="training"/>
    <connect from_op="Process Documents from Files" from_port="word list" to_op="Store wordlist" to_port="input"/>
    <connect from_op="Validation" from_port="model" to_op="Store model" to_port="input"/>
    <connect from_op="Validation" from_port="averagable 1" to_port="result 1"/>
    <portSpacing port="source_input 1" spacing="0"/>
    <portSpacing port="sink_result 1" spacing="0"/>
    <portSpacing port="sink_result 2" spacing="0"/>
  </process>
</operator>
</process>

```

## K. Sentiment analiza na osnovu leksikona: Primenjeni algoritmi

Sentiment analiza na osnovu leksikona implementirana je u Python programskom jeziku prema sledećim algoritmima:

---

### Algoritam K.1. Preprocesiranje dokumenata

---

**Input:** Skup dokumenata koji se analiziraju (cele recenzije ili pojedinačne rečenice).

**Output:** Dokumenti razdvojeni po rečenicima, gde svaka rečenica predstavlja listu reči koje je sačinjavaju.

**D** ← skup dokumenata  
**R** ← skup rečenica u dokumentu  
**T** ← skup tokena u rečenici

```

for each d ∈ D do
    razdvoj d po rečenicima (R)
    for each r ∈ R do
        razdvoj r po tokenima (T)
    return T
end for

```

---

### Algoritam K.2. Identifikovanje reči u rečnicima

---

**Input:** T ← preprocesiranje Dokumenta (d)

→ Koristi Algoritam K.1.

**Output:** Rečenice razdvojene po rečima, gde je svaka reč koja je element nekog rečnika označena adekvatnim oznakama.

**L** ← skup leksikona (rečnik pozitivnih sentiment izraza, rečnik negativnih sentiment izraza, negacije, intenzifikatori i neutralizatori)

**lr** ← reč u rečniku

**t** ∈ T

```

for each l ∈ L do
    while rečenica ima tokene do
        if len(lr) > len(t) then
            aktivan_token = lr
        else
            aktivan_token = t
        end if

        if aktivan_token ∈ l and aktivan_token ∈ T then
            označi aktivan token u rečenici pokazivačem vrste reči (pozitivna, negativna, inv, int,
                                                                    dec)
            označena_rečenica += aktivan_token sa oznakom
        else if
            označena_rečenica += aktivan_token bez oznake
        end if
    end while
    return označena_rečenica
end for

```

---

**Algoritam K.3.** Dodeljivanje vrednosti sentimenta na osnovu oznaka**Input:** Preprocesirani dokumenti ( $d$ ).**Output:** Sentiment vrednost reči u rečenici. $O \leftarrow$  označiReči ( $d$ ) $\rightarrow$  Koristi Algoritam K.2.

```

for each  $o \in O$  do
  if  $o = \text{'pozitivno'}$  then
    return 1
  else if  $o = \text{'negativno'}$  then
    return -1
  else if  $o = \text{'inv'}$  then
    return -2
  else
    return 0
  end if
end for

```

**Algoritam K.4.** Kalkulisanje sentiment skora analiziranog dokumenta**Zahtevi:** Sentiment leksikoni negativnih i pozitivnih reči, rečnik negacija i drugih reči koje obrću sentiment naredne u nizu, rečnik intenzifikatora sentimenta i rečnik neutralizatora sentimenta.**Input:** Rečenica ili recenzija  $s$  (input zavisi od nivoa klasifikacije).**Output:** Sentiment polaritet rečenice/recenzije  $s$  na osnovu izračunatog sentiment skora  $p(s)$ .

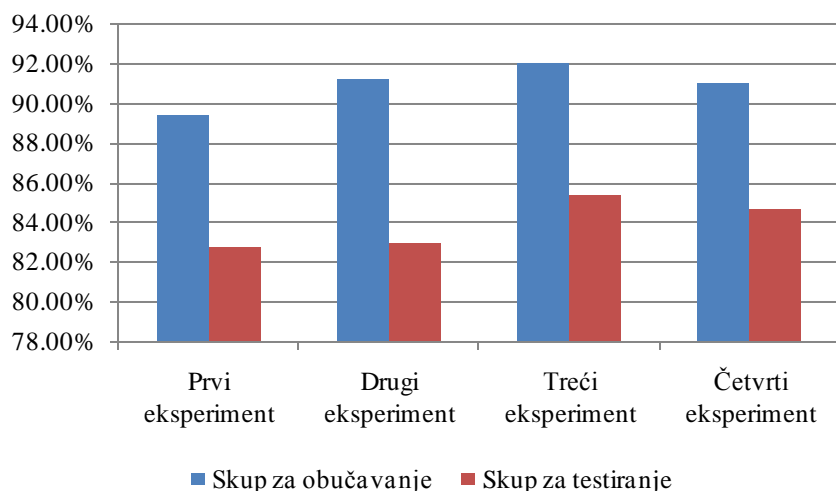
1. Primeni algoritam K.2. radi identifikovanja svih jedinstvenih podudaranja reči iz rečnika  $L$  u rečenici/recenziji  $s$ .
2. Za svaki  $l \in L$  koji je deo rečenice  $s$  kalkulisati vrednost sentiment polariteta  $p_l$  primenom algoritma K.3. i uzimajući u obzir sledeće:
  - a. Ukoliko reč iz rečnika negacije (u oznaci *inv*) prethode sentiment reči obrnuti sentiment postavljanjem  $p_l = p_l * (-1)$ .
  - b. Ukoliko reč iz rečnika intenzifikatora prethodi sentiment rečima pojačati intenzitet sentimenta postavljanjem  $p_l = p_l * 2$ .
  - c. Ukoliko reč iz rečnika neutralizatora prethodi sentiment rečima umanjiti intenzitet sentimenta postavljanjem  $p_l = p_l / 2$ .
3. Izračunati sentiment skor postavljanjem  $p(s) = \sum_{l \in L} p_l$ .
4. Definiši sentiment polaritet svakog  $s$ :
  - a. Ukoliko je  $p(s) > 0$  postavi sentiment polaritet  $s$  na pozitivan.
  - b. Ukoliko je  $p(s) < 0$  postavi sentiment polaritet  $s$  na negativan.
  - c. Ukoliko je  $p(s) = 0$  postavi sentiment polaritet  $s$  na neutralan.

## L. Evaluacione mere binarne klasifikacije sentimenta zasnovane na leksikonima

Eksperimenti, koji su opisani u [Poglavlju 4.1.6.2.](#) i realizovani na nivou dokumenta i na nivou rečenice, reprodukovani su u binarnoj klasifikaciji, koja je obuhvatila klase pozitivnih i negativnih recenzija/rečenica. Cilj realizovanih analiza je evaluacija uticaja neutralne klase, odnosno njenog izostavljanja iz klasifikacije, na performanse klasifikacije sentimenta zasnovane na leksikonima.

### Klasifikacija na nivou dokumenta

Slika L.1. ilustruje tačnost klasifikacije modela u četiri realizovana eksperimenta, iskazanu merom *Accuracy*. Na osnovu prikazanih podataka može se zaključiti da treći eksperiment daje najbolje rezultate.



Slika L.1. Tačnost modela – mera *Accuracy* – skup za obučavanje/nivo dokumenta

Prvi eksperiment: Sentiment rečnici obuhvataju sve reči koje su u postupku anotacije anotatori označili kao pozitivne ili negativne. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

Tabela L.1. Eksperiment 1: Preciznost rada modela – skup za obučavanje

Accuracy: 89.45%			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1581	122	92.83%
Predviđeno negativno	150	724	82.84%
Recall	91.33%	85.58%	
<b>F-mera</b>	<b>92.07%</b>	<b>84.19%</b>	



**Tabela L.2. Eksperiment 1: Preciznost rada modela – skup za testiranje**

<b>Accuracy: 82.77%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	597	91	86.77%
Predviđeno negativno	77	210	73.17%
Recall	88.58%	69.77%	
<b>F-mera</b>	<b>87.67%</b>	<b>71.43%</b>	

Drugi eksperiment: Sentiment rečnici su manuelno ažurirani. Uklonjeni su pojmovi koji su tokom anotacije označeni kao sentiment reči, ali ne nose sentiment, kao i pojmovi koji su pogrešnim sentimentom označeni. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

**Tabela L.3. Eksperiment 2: Preciznost rada modela – skup za obučavanje**

<b>Accuracy: 91.23%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1647	142	92.06%
Predviđeno negativno	84	704	89.34%
Recall	95.15%	83.22%	
<b>F-mera</b>	<b>93.58%</b>	<b>86.17%</b>	

**Tabela L.4. Eksperiment 2: Preciznost rada modela – skup za testiranje**

<b>Accuracy: 82.92%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	610	100	85.92%
Predviđeno negativno	65	191	74.61%
Recall	90.37%	65.64%	
<b>F-mera</b>	<b>88.09%</b>	<b>69.84%</b>	

Treći eksperiment: Upotrebljeni su manuelno korigovani sentiment rečnici, rečnik negacije i reči koje menjaju orijentaciju sentimenta naredne reči u rečenici, rečnik intenzifikatora i rečnik neutralizatora. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

**Tabela L.5. Eksperiment 3: Preciznost rada modela – skup za obučavanje**

<b>Accuracy: 92.08%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1664	137	92.39%
Predviđeno negativno	67	709	91.37%
Recall	96.13%	83.81%	
<b>F-mera</b>	<b>94.22%</b>	<b>87.43%</b>	

**Tabela L.6. Eksperiment 3: Preciznost rada modela – skup za testiranje**

<b>Accuracy: 85.35%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	617	85	87.89%
Predviđeno negativno	58	216	78.83%
Recall	91.41%	71.76%	
<b>F-mera</b>	<b>89.62%</b>	<b>75.13%</b>	

Četvrti eksperiment: Upotrebljeni su manuelno korigovani sentiment rečnici, rečnik negacije i reči koje menjaju sentiment polaritet naredne reči u rečenici, s tim da je ojačan sentiment ovih reči prema algoritmu K.3. prikazanom u [Poglavlju K](#), kao i rečnik intenzifikatora i rečnik neutralizatora. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

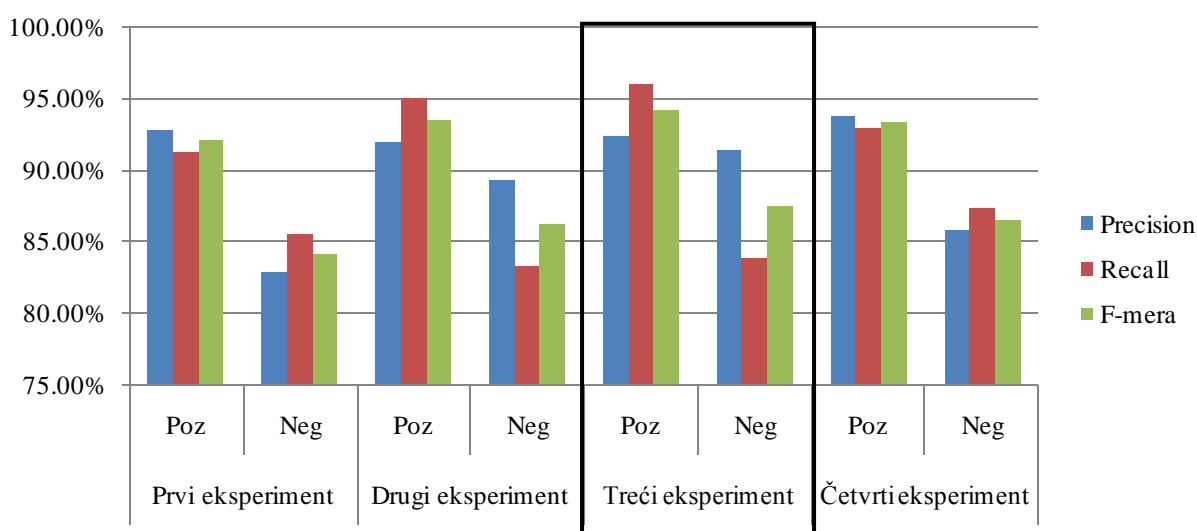
**Tabela L.7. Eksperiment 4: Preciznost rada modela – skup za obučavanje**

<b>Accuracy: 91.11%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1609	107	93.76%
Predviđeno negativno	122	739	85.83%
Recall	92.95%	87.35%	
<b>F-mera</b>	<b>93.35%</b>	<b>86.58%</b>	

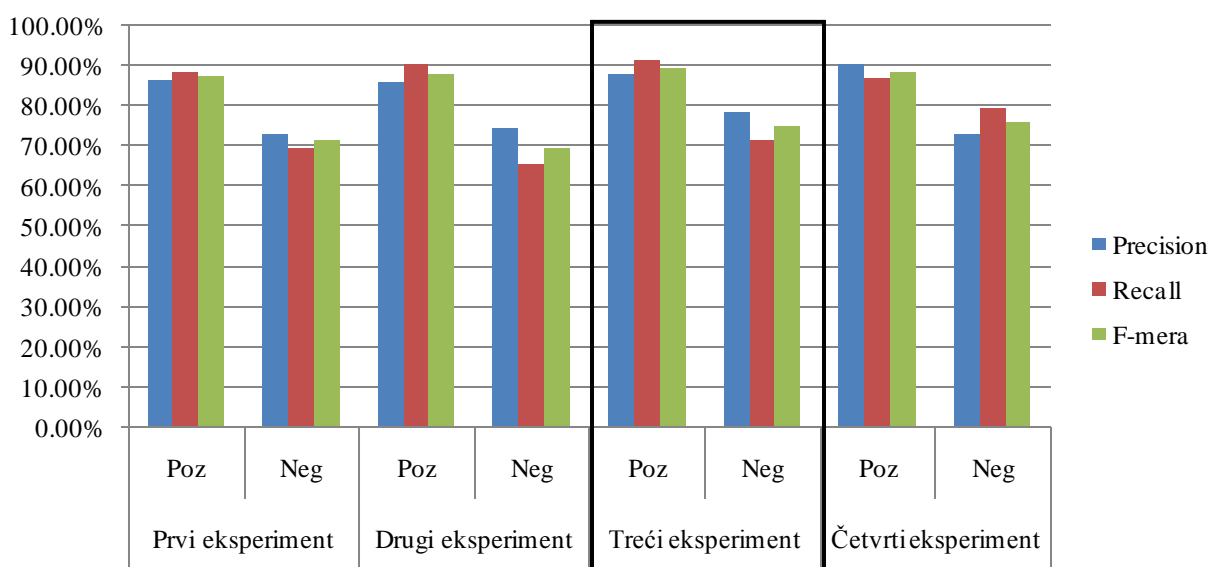
**Tabela L.8. Eksperiment 4: Preciznost rada modela – skup za testiranje**

<b>Accuracy: 84.73%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	588	62	90.46%
Predviđeno negativno	87	239	73.31%
Recall	87.11%	79.40%	
<b>F-mera</b>	<b>88.75%</b>	<b>76.23%</b>	

Naredne dve slike sumiraju preciznost rada izgrađenih modela nad podacima iz skupa za obučavanje i nad podacima iz skupa za testiranje, paralelno za sve realizovane eksperimente. Na osnovu podataka može se zaključiti da se preciznost rada modela povećava uvođenjem dodatnih rečnika (modifikatora i reči koje menjaju sentiment polaritet naredne reči u tekstu). Veće oscilacije u kretanju vrednosti evaluacionih mera zabeležene su u rezultatima rada modela nad podacima iz skupa za obučavanje, slika L.2. Kao što je već istaknuto, za primenu razvijenog modela donekle je bitnija mera *Precision*, odnosno procenat slučajeva u kojima će model tačno identifikovati određenu klasu. Ukoliko se ova mera posmatra samostalno, najbolji rezultati po pitanju negativne klase postignuti su u trećem eksperimentu: u 78.83% slučajeva model će tačno identifikovati negativnu recenziju, kao što slika L.3. pokazuje.



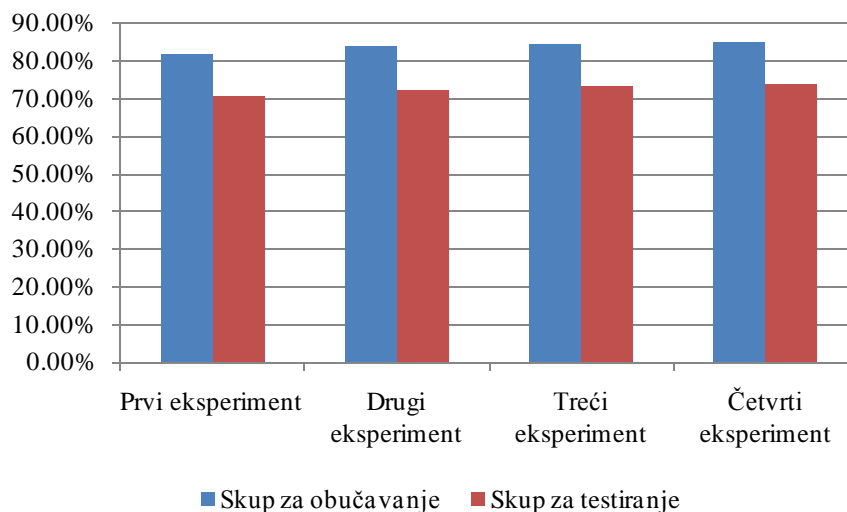
Slika L.2. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za obučavanje



Slika L.3. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za testiranje

### Klasifikacija na nivou rečenice

Slika L.4. ilustruje tačnost klasifikacije modela u četiri realizovana eksperimenta, iskazanu merom *Accuracy*. Na osnovu prikazanih podataka može se zaključiti da postoje male oscilacije u tačnosti klasifikacije, kada je ona iskazana merom *Accuracy*. U četvrtom eksperimentu su dobijene najviše vrednosti ove mere za obe klase.



Slika L.4. Tačnost modela - mera *Accuracy* – nivo rečenice

Prvi eksperiment: Upotrebljeni su manuelno korigovani sentiment rečnici. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

Tabela L.9. Eksperiment 1: Preciznost rada modela – skup za obučavanje

Accuracy: 82.08%			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	2975	490	85.86%
Predviđeno negativno	515	1627	75.96%
Recall	85.24%	76.85%	
<b>F-mera</b>	<b>85.55%</b>	<b>76.40%</b>	

Tabela L.10. Eksperiment 1: Preciznost rada modela – skup za testiranje

Accuracy: 70.78%			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1148	394	74.45%
Predviđeno negativno	303	540	64.06%
Recall	79.12%	57.82%	
<b>F-mera</b>	<b>76.71%</b>	<b>60.78%</b>	

Drugi eksperiment: Upotrebljeni su manuelno korigovani sentiment rečnici i rečnik negacije i drugih pojmova koji menjaju orijentaciju sentimenta naredne reči u rečenici. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

**Tabela L.11. Eksperiment 2: Preciznost rada modela – skup za obučavanje**

<b>Accuracy: 84.31%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	3155	545	85.27%
Predviđeno negativno	335	1572	82.43%
Recall	90.40%	74.25%	
<b>F-mera</b>	<b>87.76%</b>	<b>78.13%</b>	

**Tabela L.12. Eksperiment 2: Preciznost rada modela – skup za testiranje**

<b>Accuracy: 72.45%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1197	403	74.81%
Predviđeno negativno	254	531	67.64%
Recall	82.49%	56.85%	
<b>F-mera</b>	<b>78.46%</b>	<b>61.78%</b>	

Treći eksperiment: Upotrebljeni su manuelno korigovani sentiment rečnici, rečnik negacije i drugih pojmova koji menjaju orijentaciju sentimenta naredne reči u rečenici, rečnik intenzifikatora i rečnik neutralizatora. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

**Tabela L.13. Eksperiment 3: Preciznost rada modela – skup za obučavanje**

<b>Accuracy: 84.93%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	3162	517	85.95%
Predviđeno negativno	328	1600	82.99%
Recall	90.60%	75.58%	
<b>F-mera</b>	<b>88.21%</b>	<b>79.11%</b>	

**Tabela L.14. Eksperiment 3: Preciznost rada modela – skup za testiranje**

<b>Accuracy: 73.27%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1201	388	75.58%
Predviđeno negativno	250	548	68.67%
Recall	82.77%	58.55%	
<b>F-mera</b>	<b>79.01%</b>	<b>63.21%</b>	

Četvrti eksperiment: Upotrebljeni su manuelno korigovani sentiment rečnici, rečnik negacije i reči koje menjaju sentiment polaritet naredne reči u rečenici, s tim da je ojačan sentiment ovih reči prema algoritmu K.3. prikazanom u [Poglavlju K](#), kao i rečnik intenzifikatora i rečnik neutralizatora. Naredne tabele daju prikaz preciznosti rada modela nad podacima iz skupa za obučavanje i skupa za testiranje.

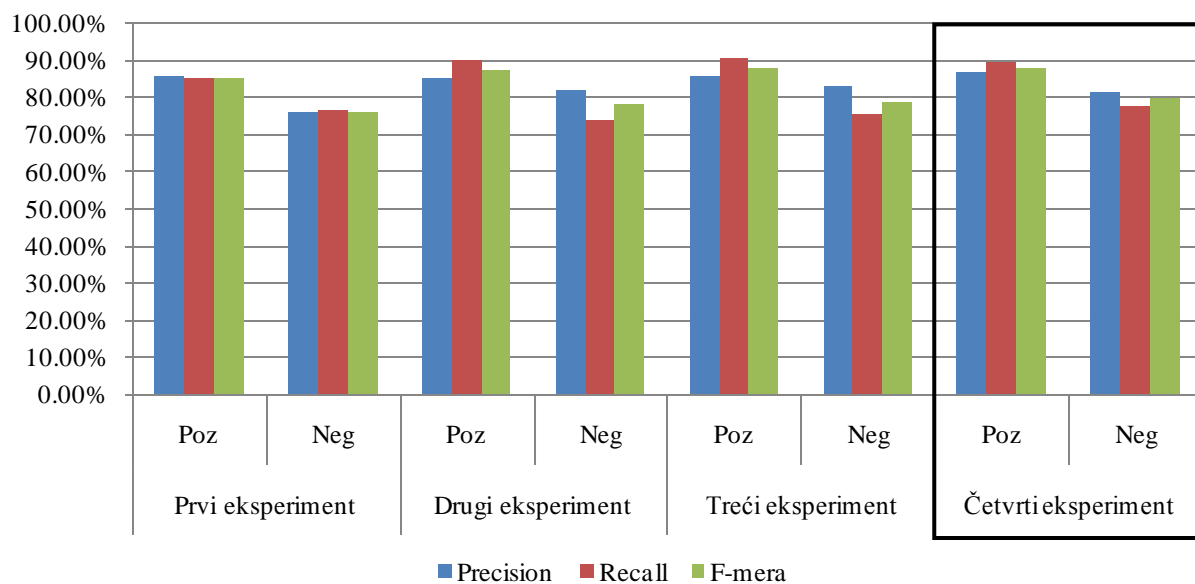
**Tabela L.15. Eksperiment 4: Preciznost rada modela – skup za obučavanje**

<b>Accuracy: 85.11%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	3125	470	86.93%
Predviđeno negativno	365	1647	81.86%
Recall	89.54%	77.80%	
<b>F-mera</b>	<b>88.22%</b>	<b>79.78%</b>	

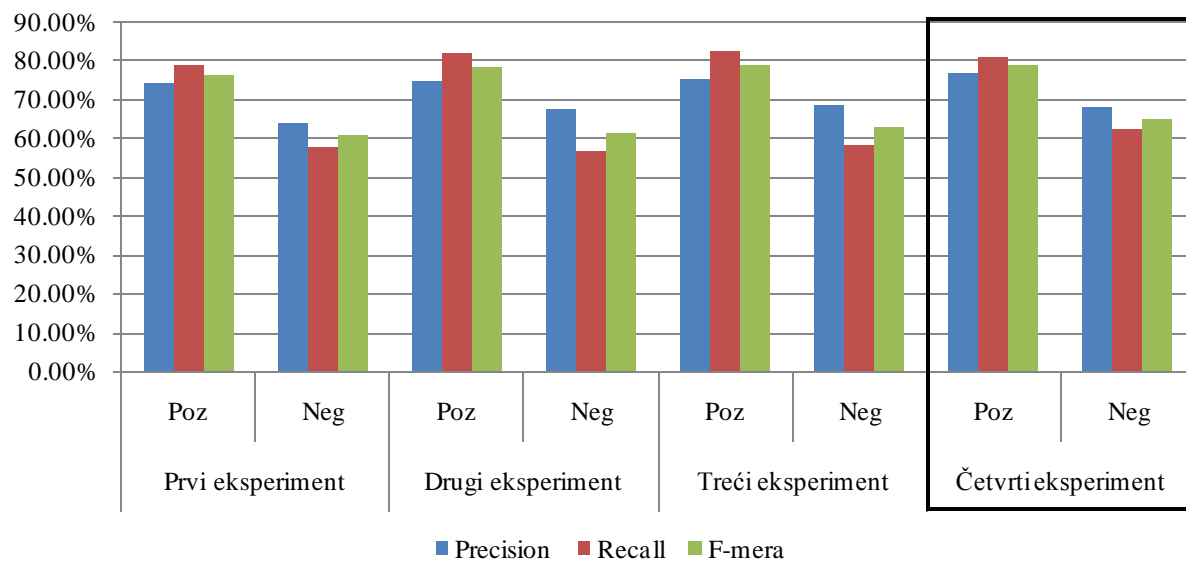
**Tabela L.16. Eksperiment 4: Preciznost rada modela – skup za testiranje**

<b>Accuracy: 73.96%</b>			
	Stvarno pozitivno	Stvarno negativno	Precision
Predviđeno pozitivno	1181	351	77.09%
Predviđeno negativno	270	583	68.35%
Recall	81.39%	62.42%	
<b>F-mera</b>	<b>79.12%</b>	<b>65.25%</b>	

Naredna dve slike prikazuju preciznost rada izgrađenih modela nad podacima iz skupa za obučavanje i nad podacima iz skupa za testiranje, paralelno za sve realizovane eksperimente. Na osnovu podataka može se zaključiti da uvođenje rečnika reči koje menjaju sentiment polaritet naredne reči u rečenici uvećava tačnost klasifikacije pozitivnih rečenica za 1.75%, dok se tačnost klasifikacije negativnih rečenica uvećava za 1%. Nad skupom za testiranje *F-mera* za negativne rečenice iznosi 61.78%, međutim, preciznost klasifikatora (*Precision*) uvećava se sa 64.06%, koliko je iznosila u prvom eksperimentu, na 67.64% u drugom eksperimentu, što govori da će u 67.64% nepoznatih slučajeva klasifikator uspešno prepoznati negativnu rečenicu. Uvođenjem rečnika pojmova koji jačaju ili umanjuju iskazani sentiment tačnost klasifikovanja negativnih rečenica podiže se na 79.11% nad skupom za obučavanje i 63.21% nad skupom za testiranje. Četvrti eksperiment, u kojem je pojačan intenzitet reči koje menjaju polaritet naredne reči, daje najbolje rezultate po pitanju identifikovanja rečenica u kojima je iskazan negativan sentiment, *F-mera* iznosi 79.78% nad skupom za obučavanje i 65.25% nad skupom za testiranje.



Slika L.5. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za obučavanje



Slika L.6. Paralelni prikaz mera preciznosti rada klasifikacije na nivou recenzija – skup za testiranje