

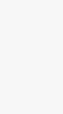
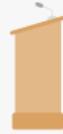
Prava frekvenca – analiza žaljivega govora v spletnih komentarjih

Urška Vranjek Oslak
Institut za slovenski jezik Franja Ramovša ZRC SAZU
urska.vranjek@zrc-sazu.si

Ajda Centa
Ljubljana
centajda@gmail.com

Žaljivi govor

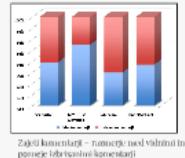
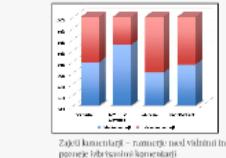
Razčlenitev definicij spornih oblik komuniciranja (sovražni, nesprejemljivi, neprimerni oz. žaljivi govor)



Uvod

- raziskovanje in diskriminacija spletnih komunikatorjev
- izboljševanje žaljivosti
- nadzor nad vsebino spletnih komentarjev
- mesto žaljivosti v spletnih komentarjih
- projekt trdne referenčnosti – temelj na sovražnih komunikacijah

- Žaljivi govor – kriterij:
 - preverjanje vseh vlog
 - izboljševanje žaljivosti
 - raziskovanje in diskriminacija
 - delo



Vsi žaljivi komentarji v enem rezultatu za raziskovanje žaljivosti

Raziskava

3 faze:

- izdelava korpusa spletnih komentarjev
- ročno razvrščanje dela komentarjev
- analiza rezultatov razvrščanja

Zajem besedil v korpus spletnih komentarjev

- komentari v vseh vlogah
- izboljševanje žaljivosti
- tehnika žaljivosti + preverjanje nad besedili tematike in vsebini komentarjev

Izvrševanje in označevanje komentarjev

- izboljševanje žaljivosti
- tehnika žaljivosti
- tehnika žaljivosti + preverjanje nad besedili tematike in vsebini komentarjev

Analiza rezultator

Komentarji	Podzemljek
– komentari v vseh vlogah	100%
– izboljševanje žaljivosti	70%
– tehnika žaljivosti	20%
– tehnika žaljivosti + preverjanje nad besedili tematike in vsebini komentarjev	0%

Sklep

- žaljivi govor → sovražni govor
- korpus spletnih komentarjev s podkorpusom ročno označenih komentarjev
- nadaljuje raziskovanje žaljivosti na spletu
- uporabnost za strojno učenje programov za samodejno zaznavanje žaljivega govorja
- boljši nadzor nad žaljivim govorom
- obvladovanje pester spletnih komunikacij



Število žaljivih	Podzemljek	Število žaljivih
– komentari v vseh vlogah	22	14
– izboljševanje žaljivosti	15	11
– tehnika žaljivosti	4	3
– tehnika žaljivosti + preverjanje nad besedili tematike in vsebini komentarjev	0	0

Nadzor nad vsebino spletnih komentarjev na spletu je pomembna kar zmanjša žaljivost

Število objavljenih	Število objavljenih
– komentari v vseh vlogah	440
– izboljševanje žaljivosti	30
– tehnika žaljivosti	10

Analiza žaljivosti na spletu je pomembna kar zmanjša žaljivost



Prava frekvanca – analiza žaljivega govora v spletnih komentarjih

Urška Vranjek Ošlak

Inštitut za slovenski jezik Fran Ramovša ZRC SAZU

urska.vranjek@zrc-sazu.si

Ajda Centa

Ljubljana

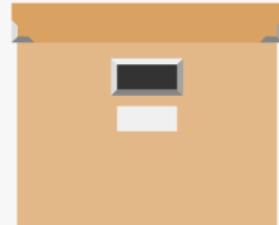
centajda@gmail.com

Uvod

- razširjenost in dostopnost spletne komunikacije
- anonimnost » **žaljivost**
- nadzor nad vsebino spletnih komentarjev

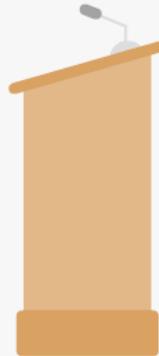
- raziskava žaljivosti v spletnih komentarjih
- projekt Prava frekvenca - korak k nenasilni komunikaciji

- žaljivi govor v literaturi
- predstavitev raziskave
 - zajem spletnih komentarjev
 - razvrščanje v kategorije
 - analiza rezultatov
- sklep



Žaljivi govor

Razčlenitev definicij spornih oblik komuniciranja (sovražni, nesprejemljivi, neprimerni oz. žaljivi govor)



Sociolingvistične raziskave

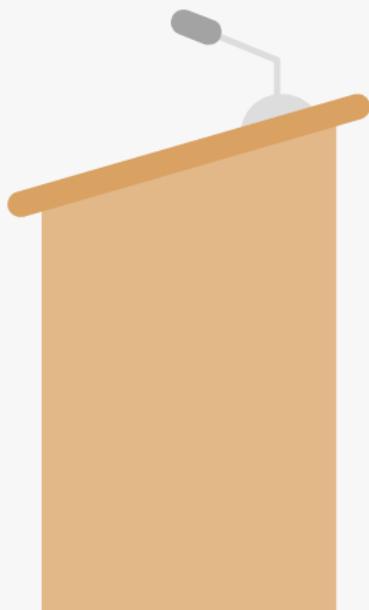
- Vinko Gorjanc (2019) – žaljivost v slovenskih leksikonskih priručnikih
- Campos Ferrera et al. (2022) – prepoznavanje elementov sovražnosti v izbranem besedilu
- Kunšek et al. (2019) – značilnosti medijskih razdaljitev

Žaljivi govor na spletnih portalih

- srečanje z moderatorji MMC RTV Slovenija in Planet Siol.net
- značilnosti moderatorskega dela:
 - število moderatorjev
 - programska uprema
 - uredniška politika

Žaljivi govor

Razčlenitev definicij spornih oblik komuniciranja (sovražni, nesprejemljivi, neprimerni oz. žaljivi govor)



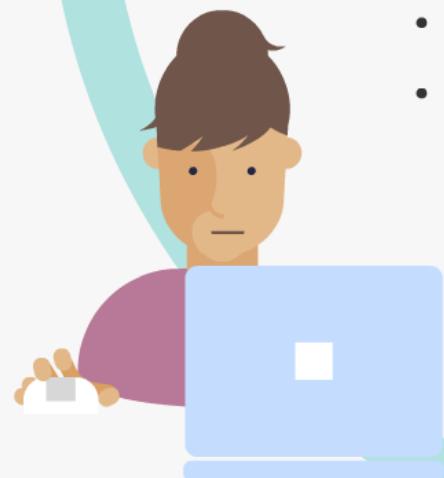
Sociolinguistične raziskave



- Vojko Gorjanc (2005) - žaljivost v slovenskih jezikovnih priročnikih
- Campos Ferreira et al. (2012) - prepoznavanje elementov sovražnosti v izbranem besedilu
- Korošec et al. (2002) - značilnosti medijskih razžalitev

Žaljivi govor na spletnih portalih

- srečanje z moderatorji MMC RTV Slovenija in Planet Siol.net
- značilnosti moderatorskega dela:
 - število moderatorjev
 - programska oprema
 - uredniška politika



a

Raziskava



3 faze:

- izdelava korpusa spletnih komentarjev
- ročno razvrščanje dela komentarjev
- analiza rezultatov razvrščanja

Zajem besedil v korpus spletnih komentarjev

- komentarji s treh portalov
- računalniško orodje iz projekta Prava frekvenca 2014 - pregled nad branostjo novic in odzivi bralcev



Analiza rezultatov

Kategorija	Odsotek komentarjev
Nepovezani komentarji	24 %
Troll	8 %
Kletvice, žaljivke, vulgarizmi	13 %
Grožnje, napadi, žaljivi govor:	55 %
- osčeni napad	19 %
- napad na skupino	59 %
- drugo	22 %

Zastopanost kategorij neprimernih komentarjev po mišnem razvrščanju



Razvrščanje in označevanje komentarjev

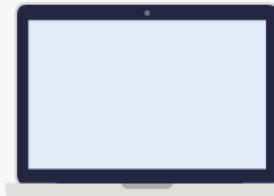
- ročno razvrščanje: 32 965 komentarjev na izbrane teme
- kategorije:
 - nepovezani komentarji
 - trolli
 - kletvice, žaljivke, vulgarizmi
 - grožnje, napadi, žaljivi govor
 - osčeni napad
 - napad na skupino
 - drugo

- žaljivi
- korpus podprt kome
- nadalj
- upora samoč
- boljši
- ohran



Zajem besedil v korpus spletnih komentarjev

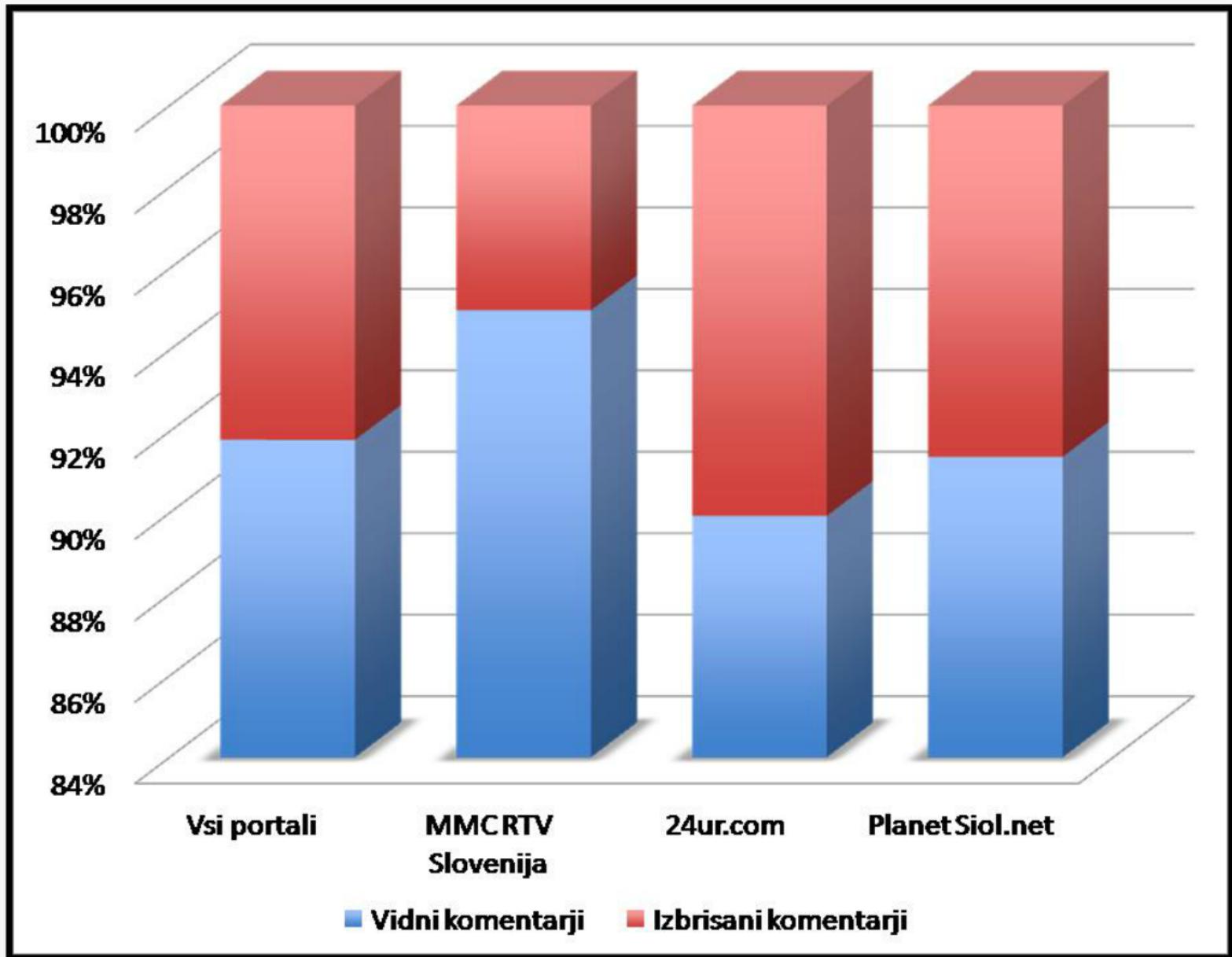
- komentarji s treh portalov
- računalniško orodje iz projekta Prava frekvenca 2014 - pregled nad branostjo novic in odzivi bralcev





Modularna zgradba računalniškega sistema

- zajem podatkov ob 4. uri zjutraj
- posodabljanje - zaznavanje izbrisanih komentarjev
- zbiranje komentarjev v bazi
- lematizacija (LemmaGen)
- obdelava s pomočjo uporabniškega vmesnika



Zajeti komentarji – razmerje med vidnimi in pozneje izbrisanimi komentarji

Razvrščanje in označevanje komentarjev

- ročno razvrščanje: 32.563 komentarjev na izbrane teme
- kategorije:
 - nepovezani komentarji
 - troli
 - kletvice, žaljivke, vulgarizmi
 - grožnje, napadi, žaljivi govor
 - osebni napad
 - napad na skupino
 - drugo

 Domov Statistika novic Ocenjevanje sovražnega
govora Moje novice v ocenjevanju Skupno ocenjevanje Statistika ocenjevanja Primerjava ocenjevanja Primerjava ocenjevanja in
brisanja Oddaja časovnic Zip & Download časovnic Forum To-do O projektu

Prava frekvenca

Ocenjevanje sovražnega govora

Napredek

0%

Novica

Ocenjevani komentar



ID novice: 4048966

 Komentar je objavljen

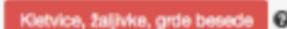
pa kaj , pozabili ste še napisati kdaj Kopitarja nekam prime !!!!

Ostali komentarji

Ocenjevalni parametri

Izberite primerno oceno komentara.

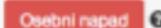
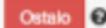
Kletvice, žaljivke, grde besede

 Kletvice, žaljivke, grde besede 

Troll

 Troll 

Grožnje, osebni napadi, žaljivi govor

 Osebni napad  Napad na skupino  Ostalo 

Random topic

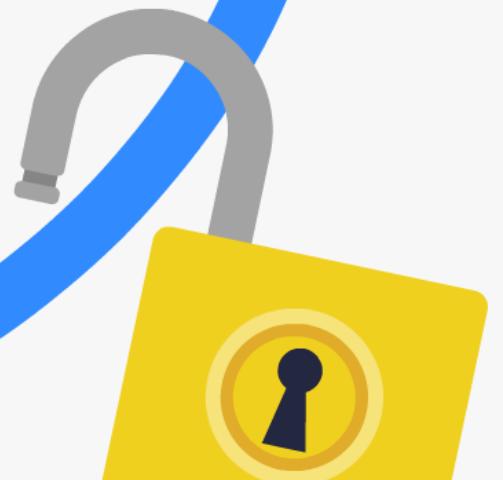
 Random topic  Karantena  Ni problematičen  Problematičen

Videz uporabniškega vmesnika za ročno razvrščanje in označevanje komentarjev

Analiza rezultatov

Kategorija	Odstotek komentarjev
Nepovezani komentarji	24 %
Trol	8 %
Kletvice, žaljivke, vulgarizmi	13 %
Grožnje, napadi, žaljivi govor:	55 %
– osebni napad	19 %
– napad na skupino	59 %
– drugo	22 %

Zastopanost kategorij neprimernih komentarjev po ročnem razvrčanju



Nepovezani komentarji:

so me zbrisali censorji fašistični, zato sem napisal še enkrat

Kar brišite. To je VAŠ kodeks. Vaš domet razumevanja.... Potem pa se čudite? A imate potem o čem pisati?

Primca zopet kar nekaj časa ne bo v službo, kjer dobiva plačo iz državne blagajne. Evo, admin, lahko zopet brišeš.

Trolanje:

Treba je spoštovati pravico staršev homoseksualcev do vnukov.



Fantek, kako se kliče tvoj očka? Marjan. Kaj pa mamica? Vinko.

Otrok = Mama + Oče Otrok je tudi mama + poštar ali pa duhovnik + deklica



Kletvice, žaljivke, vulgarizmi:
*Pa kaj tem **kretenom** ni jasno?!!*



*Do tega pride ko te začnejo **pedri** posiljevati.*

Sm pa zihr da vsi tipi, ki so tako proti homoseksualcem, drkajo na lezbične porniče...

Grožnje, napadi, žaljivi govor:

*Še nikjer pa nisem zasledil, da bi se homoseksualci komu maščevali.
Tudi to kaže kakšen **lažljivec je Primc.***

*Cerkev je proti, ker **duhovnikom ni tako zabavno in vznemrljivo z s polnoletnimi ministranti in farani moškega spola, če se legalizira.**
Hipokrati.*

*Imajo **MOŽNOST**, da se postavijo v dolgo čakalno vrsto, nič drugega
kot to! BRAVO G.Bruno, tako pač dela SDS, tudi podpise ponareja!*



	Število objavljenih oz. sprejemljivih komentarjev	Število izbrisanih oz. nesprejemljivih komentarjev
Moderatorji	410	143
Označevalka 1	495	58
Označevalka 2	483	70
Označevalka 3	456	97



Razvrstitev komentarjev iz testne množice

	Neprimerno	Žaljivi govor
Moderatorji	143	99
Strinjanje O1 + O2 + O3	27	9
Strinjanje O1 + O2	29	17
Strinjanje O1 + O3	36	10
Strinjanje O2 + O3	44	23
Strinjanje samo O1 + O2	2	8
Strinjanje samo O1 + O3	19	1
Strinjanje samo O2 + O3	17	14
Ni strinjanja	78	67

Strinjanje ocenjevalk pri odločanju o neprimernosti komentarjev in razvrščanju v kategorije žaljivosti

Sklep

- žaljivi govor » sovražni govor
- **korpus spletnih komentarjev s podkorpusom ročno označenih komentarjev**
- nadaljnje raziskovanje žaljivosti na spletu
- uporabnost za strojno učenje programov za samodejno zaznavanje žaljivega govora
- boljši nadzor nad žaljivim govorom
- ohranjanje pestre spletne komunikacije

Prava frekvenca – analiza žaljivega govora v spletnih komentarjih

Urška Vranjek Oslak
Institut za slovenski jezik Franja Ramovša ZRC SAZU
urska.vranjek@zrc-sazu.si

Ajda Centa
Ljubljana
centajda@gmail.com



	Besedilo objavljeno na spletnih komentarjih	Besedilo objavljeno na nežaljivih komentarjih
Mestobring	0,1%	11,1%
Domestika	40,1%	50,0%
Domestika 2	59,9%	0,0%

Korpus je razvrščen na 4 kategorije: žaljivo, nežaljivo, žaljivo nežaljivo, nežaljivo žaljivo.

