

Računalniška analiza čustev in tem v Wikiviru

DOI: 10.4312/Obdobja.44.269-278

Damjan Popič, Filozofska fakulteta, Univerza v Ljubljani, Ljubljana; damjan.popic@ff.uni-lj.si

V prispevku predstavimo računalniško analizo tem in čustev v korpusu slovenske književnosti, prosto dostopni v repozitoriju Wikivir. Pri tem predstavimo tako pristop k izgradnji korpusa, ki ga želimo v analizo ponuditi tudi drugim raziskovalcem v okviru platforme SketchEngine in na repozitoriju Clarin, kot tudi metodo za analizo čustev in prevladujočih tematik v obsežnih podatkovnih zbirkah. Pri raziskavi poskušamo izpostaviti glavne tematske poudarke v različnih časovnih obdobjih, žanrih in pri različnih avtorjih ter prevladujoča čustva, pri čemer uporabimo tako leksikonski pristop kot tudi analizo sentimenta. Kljub neprecenljivemu bogastvu digitaliziranega slovenskega leposlovja na Wikiviru je bil ta vir zaradi zapletenosti označevalnika MediaWiki, razpršenosti prenesenih besedil ter zahtevnosti pretvarjanja in procesiranja besedil do zdaj slabo izkoriščen za obsežnejše raziskave, ki bi temeljile na računalniškem procesiranju naravnega jezika. Za namene te raziskave smo pripravili korpus prosto dostopne slovenske književnosti, ki zajema nekaj več kot 62 milijonov besed, nabranih iz 22.919 različnih besedil, ki so v Wikiviru (neconsistentno) označena z več kot 2000 različnimi kategorijami oz. metapodatki (avtor, leto, stoletje, žanr ipd.). Te podatke poskušamo z uporabo računalniških pristopov tudi sistematizirati ter natančneje predstaviti distribucijo besedil, objavljenih v Wikiviru.

procesiranje naravnega jezika, Wikivir, slovenska književnost, LDA, leksikon čustev

This article presents a computational analysis of themes and emotions in the corpus of Slovenian literature, freely available on Wikivir (i.e., the Slovenian Wikisource). It presents the approach used to build the corpus, which is also made freely available to other researchers, as well as the method for analyzing emotions and predominant topics in large databases. The research highlights the main thematic emphases in different time periods, genres, and authors, as well as the predominant emotions, using both a lexical approach and sentiment analysis. Despite the invaluable wealth of the digitized Slovenian literature on Wikivir, this source has been underutilized for more extensive research based on natural language processing due to the complexity of the MediaWiki tagger, the dispersion of transferred texts, and the complexity of text conversion and processing. For the purposes of this research, a corpus of freely accessible Slovenian literature was compiled, comprising just over 62 million words collected from 22,919 texts, which are (inconsistently) annotated on Wikivir with over two thousand categories or metadata (author, year, century, genre, etc.). These data are also being systematized using automatized approaches, thus presenting a more accurate distribution of texts published on Wikivir.

natural language processing, Wikisource, Wikivir, Slovenian literature, LDA, emotion lexicon

Alojzija Zupan Sosič (ur.): *Čustva in slovenska književnost*.
Ljubljana: Založba Univerze v Ljubljani, 2025 (zbirka Obdobja, 44)
ISSN: 1408-211X, e-ISSN: 2784-7152
ISBN 978-961-297-712-2, 978-961-297-709-2 (PDF)



1 Uvod

Prosto in odprto dostopne zbirke, kot je Wikivir,¹ imajo izjemno pomembno vlogo za ohranjanje in proučevanje kulturne dediščine, kar je eden glavnih postulatov digitalne humanistike, ki združuje različne humanistične discipline in (digitalno) tehnologijo. V tej raziskavi želimo uporabiti računalniško analizo besedil za sistematično proučevanje tematskih in čustvenih obrisov slovenske literature, s poudarkom na dveh raziskovalnih oseh: 1) ugotavljanju prevladujočih tematik in 2) mapiranju čustvenih tokov v različnih zgodovinskih obdobjih, pri različnih avtorjih ipd.

Tovrstna analiza slovenskega leposlovja še ni bila izvedena, opravljena pa je že bila primerjalna analiza tematik v slovenskih korpusih (Logar Berginc, Ljubešić 2013): primerjana sta bila korpusa Gigafida in slWac, s čimer je bilo inherentno primerjano tudi vsebovano leposlovno gradivo. Računalniški pristopi, kot sta tematsko modeliranje in analiza čustev, nam omogočajo analizo velikih zbirk besedil tudi za preverjanje hipotez, ki so izhodiščno izrazito humanistične.

S svojo raziskavo želimo napraviti uvodne korake k celostni računalniški analizi glavnih motivov v slovenski književnosti in jih vizualizirati, s čimer pa želimo predvsem a) aplicirati računalniško jezikoslovje tudi na polju slovenske književnosti in rezultate sopostaviti kvalitativnim študijam, hkrati pa tudi omogočiti nadaljnje kvalitativne raziskave, b) ustvariti vir, s katerim bo mogoče odkrivati morda še neodkrita vzorce in povezave v slovenski književnosti, c) zajeti in uporabiti metapodatke o besedilih, s katerimi bo mogoče napraviti globljo analizo, in č) zgraditi nov javno dostopen jezikovni vir za analizo slovenščine v leposlovju, s čimer raziskava tako izrablja kot tudi spodbuja potencial odprtih digitalnih repozitorijev za omogočanje naprednih znanstvenih raziskav na področju korpusnega jezikoslovja in digitalne humanistike.

2 Podatki in priprava korpusa

Kot vir podatkov smo uporabili podatkovno bazo slovenskega Wikivira, in sicer smo pridobili t. i. posnetek stanja (angl. *dump*) z dne 23. 1. 2025.² Wikivir je eno od spletišč fundacije Wikimedia, katerega glavni namen je sistematično zbiranje in hramba besedil, ki niso več avtorsko zaščitena – kar preprosto pomeni, da je od smrti avtorjev in avtoric minilo že 70 let ali pa da so se nosilci pravic tem odpovedali. Wikivir je bržkone bistveno manj poznan in obiskan portal kot sestrška Wikipedija, vendar je slovenska različica portala izjemno dejavna in obstaja že od leta 2006, tri leta manj od angleške (Hladnik 2022: 17). Brez dvoma lahko predpostavimo, da ima glavne zasluge za dejavnost slovenskega Wikivira Oddelek za slovenistiko na Filozofski fakulteti v Ljubljani, kjer je digitalizacija besedil vključena tudi v pedagoški proces, za kar so wikiji tudi zelo pripravljeni (Hladnik, Polajnar 2016),³ obenem pa je bila digitalizacija besedil v slovenščini izvedena tudi v okviru projektov (Hladnik 2008).

1 https://sl.wikisource.org/wiki/Glavna_stran

2 <https://dumps.wikimedia.org/slwikisource/>

3 Zlasti v okviru pedagoškega dela nekdanjega profesorja Mirana Hladnika.

Zapis Wikivira v formatu XML je sicer celovit in bogat z informacijami, a sam po sebi precej neprimeren za računalniško analizo, saj so besedila zapisana nestrukturirano, in kot je videti, v časovnem zaporedju dodajanja besedil,⁴ obenem pa so v zapisu zajete tudi diskusije uporabnikov, komentarji in segmenti z računalniško kodo. Zato smo med analizo izvedli številne korake, ki so temeljili na predhodnem in vmesnem kvalitativnem preverjanju korpusa:

- samodejni prepis vseh »strani« v besedila ter samodejno luščenje pripisanih kategorij, ki so v Wikiviru pripisane zelo nesistematično (denimo »dela leta 2006«, »19. stoletje« ipd.), obenem pa je tudi veliko tipkarskih napak in drugih nedoslednosti;
- odstranjevanje besedil, ki jih v to raziskavo in prihodnje ne želimo vključiti, bodisi ročno na podlagi kvalitativnega pregleda bodisi na podlagi kategorij, pripisanih besedilom, ki so omogočile, da smo lahko besedila odstranili samodejno (npr. »jugoslovanske ustave«);
- preliminarna normalizacija in sistematizacija oznak;
- dodajanje oznak, kjer jih v izvornem zapisu ni ali pa jih med postopkom nismo (pravilno) zajeli.

Izvorni zapis besedila v formatu XML in končni zapis v formatu, ki je primeren za uvoz v SketchEngine, podaja Tabela 1, gre pa za Kosovelov Kons: ABC.

4 Prvo besedilo Wikivira v zapisu XML je *Zedinjena Slovenija*, besedilo pa je bilo dodano 8. junija 2006. Kljub nesistematični obliki zapisa pa nam tovrstni, kronološki posnetek stanja Wikivira vendarle dovoljuje, da lahko korpus sistematično dopolnjujemo z novimi besedili in tako napravimo dinamični korpus.

Izvorni zapis	Končni zapis
<pre> <page> <title>Kons: ABC</title> <ns>0</ns> <id>1600</id> <revision> <id>64631</id> <parentid>16102</parentid> <timestamp>2011-06-30T07:08:10Z</ timestamp> <contributor> <username>Mihas-bot</username> <id>29</id> </contributor> <minor /> <comment>javna last</comment> <origin>64631</origin> <model>wikitext</model> <format>text/x-wiki</format> <text bytes="540" sha1="5j7ff6nj- dd7xncw2jtnsi8cf9r3oxdn" xml:space="pre- serve">{{naslov prejšnji=Evakuacija duha naslednji=Prostituirana kultura naslov=Kons: ABC poglavje=[[Integrali ,26]] avtor= Srečko Kosovel opombe= licenca=javna last }} Ostani mrzlo, srce!&lt;br/&gt; Cinik.&lt;br/&gt; Transformator.&lt;br/&gt; Orient ekspres v Parizu na viaduktu.&l- t;br/&gt; Okovi na rokah.&lt;br/&gt; Avtomobili tečejo.&lt;br/&gt; Jaz ne morem.&lt;br/&gt; Moja misel-elektrika&lt;br/&gt; je v Parizu.&lt;br/&gt; Vonj medicín&lt;br/&gt; s klinik.&lt;br/&gt; Fuj –&lt;br/&gt; Pljuj, zaničuj.&lt;br/&gt; Fuj, fuj,&lt;br/&gt; fuj!&lt;br/&gt; [[Kategorija:Integrali ,26]] [[Kategorija:Pesmi (Kosovel)]]</text> <sha1>5j7ff6njdd7xncw2jtn- si8cf9r3oxdn</sha1> </revision> </page> </pre>	<pre> <doc title="Kons: ABC" author="Srečko Kosovel" categories="integrali_26, kosovel, pesmi"> Ostani mrzlo, srce! Cinik. Transformator. Orient ekspres v Parizu na viaduktu. Okovi na rokah. Avtomobili tečejo. Jaz ne morem. Moja misel-elektrika je v Parizu. Vonj medicín s klinik. Fuj – Pljuj, zaničuj. Fuj, fuj, fuj! </doc> </pre>

Tabela 1: Primerjava začetnega in končnega zapisa besedila v Wikiviru.

Kot lahko vidimo, je izvorni zapis močno členjen, kategorije pa niso zapisane na način, ki bi bil strojno zlahka berljiv in luščljiv, zato smo s serijo samodejnih ukazov in skript poskušali spraviti vsa besedila v končno obliko. Ker je izvorni format pogosto

arbitraren, nam to ni uspelo prav povsod; besedila so sicer ohranjena v celoti, kategorije pa so v določenih primerih ohranjene še nesistematično, denimo tako, da so vsi opisi zajeti v atribut `categories` (recimo `categories=' 'kosovel, integrali_26, pesmi' '`), kar je z vidika ohranjanja podatkov sicer ustrezno, saj so ohranjeni vsi podatki, vendar niso pripisani ustreznim kategorijam, ki bi dovoljevale optimalno filtriranje.⁵ Pri nekaterih primerih smo med pripravo korpusa izgubili celo vrsto metapodatkov, kot kaže spodnji primer:

```
<doc title="Levo očetovo oko joče, desno pa se smeje">
```

Kot lahko vidimo, je v glavi besedila ostal zgolj naslov, tako da smo v takih primerih ročno dodali vsaj podatek o avtorju:

```
<doc title="Levo očetovo oko joče, desno pa se smeje"
author="Matija Valjavec">
```

Da bi bila analiza čim bolj pomenljiva, smo za različne vrste računalniške analize uporabili zgolj besedila, pri katerih so metapodatki vsaj v določeni meri ustrezno pripisani. To smo poskušali napraviti z uporabo lastnega spletnega pajka (angl. *crawler*), s katerim smo samodejno prevzeli vse kategorije s spletišča prek dostopne točke (API-ja) portala MediaWiki.

2.1 Korpus Wikivir 0.1 v številkah

Končna oblika korpusa ob pripravi prispevka zajema nekaj več kot 62 milijonov besed, kot prikazuje Tabela 2.

Parameter	Frekvenca
pojavnice	78.284.028
besede	62.641.170
povedi	4.646.656
dokumenti	22.909

Tabela 2: Korpus WikivirSLO v številkah.

Iz prvotne različice korpusa s šumom in neželenimi besedili smo zbrisali okoli tri milijone besed, treba pa je omeniti, da se številke v Tabeli 2 ne ujemajo z navedbami, ki jih podaja Wikivir na naslovni strani in v razdelku Statistika.⁶ Pri pripravi korpusa smo za besedila poleg naslova poskušali zajeti še metapodatke o avtorjih, letnici, žanru in stoletju, kategorije in njihove frekvence pa podajamo v naslednjem odseku.

2.2 Kategorije v korpusu Wikivir 0.1

V tem razdelku predstavljamo kategorije glede na metapodatke, pripisane v korpusu, obenem pa tudi po 10 najpogostejših predstavnikov. Tabela 3 zajema najpogostejše zastopane avtorje.

⁵ Korpus bomo zato poskušali z dodatnimi koraki še dopolniti, nato pa periodično dodajati besedila, ki bodo na novo objavljena na Wikiviru.

⁶ <https://sl.wikisource.org/wiki/Posebno:Statistika>

Avtor/avtorica	Frekvenca
Vera Albreht	453
Ivan Cankar	210
Miran Jarc	202
Anton Medved	158
Lojze Grozde	131
Dragotin Kette	114
Rado Murnik	96
Rudolf Maister	88
Vladimir Levstik	76

Tabela 3: Najpogostejši avtorji in frekvenca.

Kot lahko vidimo, je v korpusu kot avtorica najpogostejše pripisana Vera Albreht, vendar to ne pomeni (nujno), da je Vera Albreht tudi dejansko avtorica največ besedil – zgolj to, da je pri besedilih ta metapodatek najpogostejši, to pa seveda velja tudi za druge podatke. Tabela 4 ponazarja število besedil po stoletjih in letih.

Leto	Frekvenca	Stoletje	Frekvenca
1873	231	20.	3241
1921	182	19.	2912
1904	152	21.	142
1910	126	16.	49
1902	110	18.	14
1848	110	17.	1
1930	110		
1907	107		
1919	101		
1908	101		

Tabela 4: Besedila po letnici in stoletju objave.

Kot lahko vidimo, ima največ besedil pripisanih letnico 1873 in podatek, da so bila objavljena v 20. stoletju. Druge metapodatke bomo morali še urediti in potencialno klasificirati, tako da bomo lahko analizirali besedila tudi po žanrih, ki so v trenutni obliki korpusa pripisani nekonsistentno.

3 Analiza

V tem poglavju predstavimo preliminarno analizo korpusa Wikivir 0.1 na dveh ravneh, in sicer predstavimo analizo prevladujočih tem v celotnem korpusu in analizo čustev. Za analizo smo uporabili pristop LDA (angl. *latent dirichlet allocation*)⁷ in leksikonski pristop k analizi čustev, v prihodnosti pa nameravamo tematsko modeliranje razširiti z uporabo transformerskih modelov, kakršen je BERTopic.⁸ Princip LDA modelira korpus kot mešanico latentnih tem, medtem ko je vsako besedilo obravnavano kot mešanica teh tem. Model sam uravnava redkost tem in raznolikost dokumentov, njegova

⁷ Za analizo smo uporabili odprtokodno knjižnico GenSim (<https://radimrehurek.com/gensim/>).

⁸ <https://maartengr.github.io/BERTopic/index.html>

verjetnostna zasnova pa omogoča sklepanje o prevladujočih temah. Vse od svojega nastanka je pristop LDA zelo pogosto uporabljena metoda nenadzorovanega učenja za odkrivanje tematskih struktur v velikih zbirkah besedil (Blei 2012).

Pristop LDA smo dopolnili z leksikonsko analizo čustev, ki beleži, kako pogosto se besede, povezane s posameznimi čustvi (npr. veselje, jeza) glede na Plutchikovo (1980, 2001) segmentacijo oz. t. i. kolo čustev, pojavljajo v vsakem besedilu. Ta segment analize temelji na delu Mohammada in Turneyja (2011), ki sta s svojim leksikonom NRC⁹ ugotavljala razlike med različnimi tipi pisem med avtoricami in avtorji. Leksikon je bil preveden tudi v slovenščino, hrvaščino in nizozemščino (Daelemans idr. 2020), čustvo posamezne besede pa kodira glede na binarno vrednost (pozitivno, negativno) in njeno čustveno povezavo (jeza, pričakovanje, gnus, strah, veselje, žalost, presenečenje, zaupanje) za hrvaški, nizozemski in slovenski jezik z binarno shemo, saj vsakemu od čustev pripiše vrednost 0 ali 1. Za slovenščino je v leksikonu vsebovanih in označenih 14.182 iztočnic, kot primer pa podajamo vrstico za besedo *ljubezen*.

Zaupanje	SL	Pozitivno	Negativno	Jeza	Pričakovanje	Gnus	Strah	Veselje	Žalost	Presenečenje
1	ljubezen	1	0	0	0	0	0	1	0	0

Tabela 5: Izpis iz leksikona LiLaH.

Kot lahko vidimo v Tabeli 5, je beseda *ljubezen* označena kot pozitivna, od osmih čustev pa ima binarno vrednost 1 pripisano pri zaupanju in veselju.

3.1 Analiza prevladujočih tem

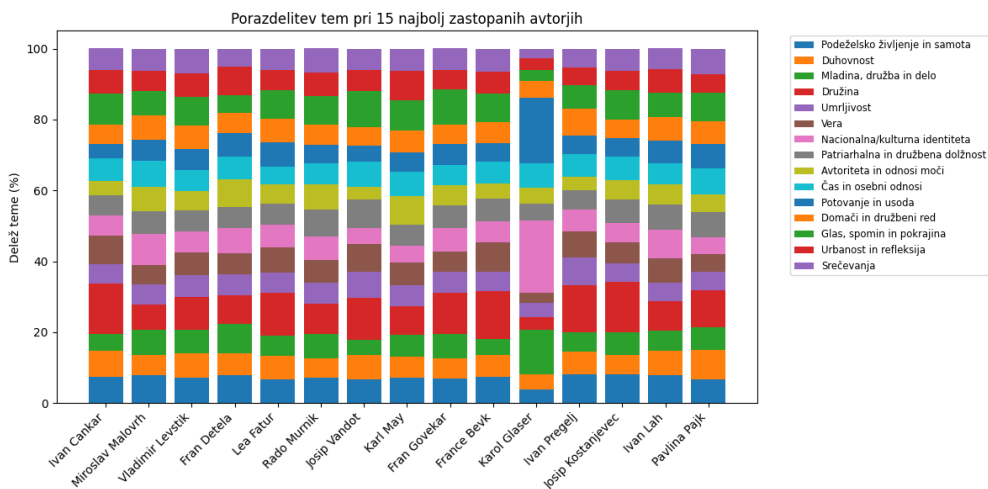
Za analizo prevladujočih tem(atik) smo upoštevali zgolj samostalnike in pridevnike, in sicer tako, da smo iz celotnega korpusa izluščili vse ter število pričakovanih tem nastavili na 15, zgornja omejitev za pojavnost posameznih besed pa je 30 odstotkov besedil (da se s tem izognemo povsem splošnemu besedišču). Izbranih 15 sklopov podaja Tabela 6, tem pa smo pripisali arbitrarne oznake za temo, da jih lahko razlikujemo, hkrati pa jih lahko tako tudi prikažemo na grafih.

9 <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

#	Najpogostejše pojavnice	Tema/semantično polje
1	drug, sam, noč, življenje, hiša, konj, dan, star	Podeželje, samota
2	srce, bog, zemlja, denar, pesem, nebo, vino, sonce	Duhovnost, čustva
3	leto, dober, mlad, delo, miza, mali, družba	Mladina, družba, delo
4	oči, beseda, mati, glava, mož, soba, vrata, grad	Družina
5	tla, cerkev, smrt, boj, letnik, nov, slab	Minljivost
6	vesel, okno, težek, luč, telo, kruh, sanje, vera	Vera
7	lep, slovenski, zadnji, cesta, jezik, janez, šola, strah	Nacionalna/kulturna identiteta
8	oče, dolg, ljubezen, volja, mir, kmet, ubog, gospodična	Patriarhalna in družbena vloga
9	gospod, stran, noga, mogoč, sila, gospodar, pravica, tovariš	Avtoriteta in odnosi moči
10	roka, čas, ura, sin, prijatelj, ime, kraj, peter	Čas in osebni odnosi
11	pot, otrok, pravi, konec, sreča, dekle, ivan, voda	Potovanje
12	svet, prvi, žena, brat, dom, večer, čuden, moč	Dom
13	glas, gospa, duša, gora, visok, pismo, korak, spomin	Glas, spomin in pokrajina
14	človek, mesto, misel, hud, stvar, srečen, reč, prazen	Urbanost, refleksija
15	dan, velik, obraz, ženska, fant, gost, vojak, kratek	Podoba

Tabela 6: 15 tematskih sklopov korpusa Wikivir 0.1.

Kot ponazarja Tabela 6, so nekateri sklopi izrazito koherentni, medtem ko so nekatere kategorije bolj heterogene, kar bomo najverjetneje lahko odpravili z uporabo transformerskih modelov – gl. npr. pojavnico *denar* pri duhovnosti ipd. Slika 1 prikazuje distribucijo identificiranih tematik pri najpogostejše zastopanih avtorjih v korpusu.



Slika 1: Distribucija tematskih polj s posameznimi avtorji in avtoricami.

Slika 1 predstavlja delež vsake od 15 kategorij pri posameznih avtorjih in avtoricah, tako da lahko vidimo, v kolikšni meri je leksika, vezana na katero od 15 pripisanih tem, zastopana pri posameznem avtorju. Opazimo lahko denimo večjo osredotočenost na sklop družina pri Cankarju, visok odstotek sklopa nacionalna/kulturna identiteta pri Glaserju zaradi narave njegovih del ipd.

3.2 Analiza prevladujočih čustev

Analiza prevladujočih čustev je pokazala rahlo nagnjenost k pozitivnim čustvom (333.479 pojavnic več nosi pozitivno oznako kot negativno). Porazdelitev po čustvih prikazuje Tabela 7.

Čustvo	Število pojavnic	Delež
Pozitivno	3.649.104	21,0 %
Negativno	3.315.625	19,0 %
Zaupanje	2.027.259	11,6 %
Pričakovanje	1.895.734	10,9 %
Veselje	1.474.383	8,5 %
Žalost	1.255.442	7,2 %
Strah	1.172.023	6,7 %
Presenečenje	947.730	5,4 %
Jeza	975.216	5,6 %
Gnus	701.088	4,0 %
Skupaj	17.413.604	100 %

Tabela 7: Porazdelitev čustvenih oznak v korpusu Wikivir 0.1.

Kot lahko vidimo, ima okoli 17,5 milijona pojavnic pripisano čustveno vrednost v leksikonu, dominantni pozitivni čustvi pa sta zaupanje (11,6 %) in pričakovanje (10,9 %), medtem ko veselje zajema 8,5 % pojavnic. Na negativnem spektru je najpogostejše čustvo žalost (7,2 %), tej pa sledijo strah (6,7 %), jeza (5,6 %) in gnus (4,0 %). Razmerje med pozitivno in negativno označeno leksiko je torej $\approx 1,10 : 1$.

4 Sklep

V prispevku smo predstavili izdelavo in strukturo korpusa Wikivir 0.1 in poskušali začetni pot računalniške obdelave korpusa, s katerim želimo v prihodnosti ustvariti dinamičen jezikovni vir z vsemi prosto dostopnimi (leposlovnimi) besedili v slovenščini. Izvedli smo preliminarno analizo tem in čustev; pri tematskem modeliranju bi lahko 15 prepoznanih tematskih polj razporedili v tri širše skupine. Prva zajema osebno os (družina, dom, čas in osebni odnosi), ki vključuje 14 % proze Ivana Cankarja, 13–14 % proze Bevka in Vandota ter znaten delež del Zofke Kveder. Druga, navzven usmerjena skupina se nanaša na kolektivno identiteto in družbene vloge (nacionalna/kulturna identiteta, patriarhalna vloga, avtoriteta in moč) in je najbolj zastopana pri avtorjih, kot sta Anton Trstenjak (26 %) ali Karol Glaser (20 %). Nazadnje je pristop LDA izpostavil krajevno dialektiko: podeželje in samota sta najpogostejše zastopana pri Trdini in Murniku, medtem ko se urbanost in refleksija najpogostejše pojavljata pri Miranu Jarcu in Vladimirju Bartolu. Vse tri tematske sklope prepletata dve eksistencialni področji: Dalmatin ima največjo težo v kategoriji vera, minljivost pa v največji meri najdemo pri Ciglerju, Jurčiču in Maistru.

Pozitivna čustvena naravnost in čustvo zaupanje skupaj obsegata tretjino od skupno 17,4 milijona čustveno označenih pojavnic (21 % oz. 11,6 %). Ta leksika je najpogostejše povezana s tematiko doma. Pričakovanje (10,9 %) je povezano s temama

potovanje in mladina, družba, delo. Negativna polariteta zajema sicer nezanemarljivih 19 %, vendar ima heterogeno distribucijo: strah, jeza in gnus skupaj zajemajo zgolj 16 % pojavnih in se združujejo pri istih avtorjih, pri katerih so pogoste teme avtoriteta, patriarhalna ali družbena vloga in minljivost.

Literatura

- BLEI, David M., 2012: Probabilistic Topic Models. *Communications of the ACM* LV/4. 77–84. <https://doi.org/10.1145/2133806.2133826>
- DAELEMANS, Walter, FIŠER, Darja, FRANZA, Jasmin, KRANJČIĆ, Denis, LEMMENS, Jens, LJUBEŠIĆ, Nikola, MARKOV, Ilia, POPIČ, Damjan, 2020: *The LiLaH Emotion Lexicon of Croatian, Dutch and Slovene*. Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1318>
- HLADNIK, Miran, 2008: Kako selimo slovensko književnost na splet. Mateja Pezdirc Bartol (ur.): *Slovenski jezik, literatura, kultura in mediji. 44. seminar slovenskega jezika, literature in kulture*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete.
- HLADNIK, Miran, POLAJNAR, Jernej, 2016: Wikiji v izobraževanju: po desetih letih izkušenj. *Andragoška spoznanja* XX/4. 73–83. <http://dx.doi.org/10.4312/as.22.4.73-83>
- HLADNIK, Miran, 2022: Wikivirovo berilo. Alenka Žbogar (ur.): *Branje v slovenskem jeziku, literaturi in kulturi. 58. seminar slovenskega jezika, literature in kulture*. Ljubljana: Založba Univerze v Ljubljani. 17–24. <https://doi.org/10.4312/SSJLK.58.17-24>
- LOGAR, Nataša, LJUBEŠIĆ, Nikola, 2013: Gigafida in slWaC: tematska primerjava. *Slovenščina* 2.0 I/1. 78–110. <https://doi.org/10.4312/slo2.0.2013.1.78-110>
- MOHAMMAD, Saif, YANG, Tony, 2011: Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*. Portland, Oregon: Association for Computational Linguistics. 70–79. <https://aclanthology.org/W11-1709.pdf>
- PLUTCHIK, Robert, 2001: The Nature of Emotions: Human Emotions Have Deep Evolutionary Roots, a Fact That May Explain Their Complexity and Provide Tools for Clinical Practice. *American Scientist* LXXXIX/4. 344–350. <http://www.jstor.org/stable/27857503>