

Zbornik konference
Jezikovne tehnologije in digitalna humanistika

Proceedings of the conference on
Language Technologies & Digital Humanities

20. september – 21. september 2018
Fakulteta za elektrotehniko, Univerza v Ljubljani
Ljubljana, Slovenija

September 20th – 21st 2018
Faculty of Electrical Engineering, University of Ljubljana
Ljubljana, Slovenia

Uredila / Edited by:
Darja Fišer, Andrej Pančur

**ZBORNİK KONFERENCE
JEZIKOVNE TEHNOLOGIJE IN DIGITALNA HUMANISTIKA**

***PROCEEDINGS OF THE CONFERENCE ON
LANGUAGE TECHNOLOGIES & DIGITAL HUMANITIES***

Uredila / Edited by: Darja Fišer, Andrej Pančur

Tehni.ni uredniki / Technical editors: Jakob Lenardič, Mihael Ojsteršek

Založil / Published by:

Znanstvena založba Filozofske fakultete v Ljubljani / *Ljubljana University Press, Faculty of Arts*

Izdal / Issued by:

Oddelek za prevajalstvo / *Department of Translation Studies*

Inštitut za novejšo zgodovino / *Institute of Contemporary History*

Za založbo / For the publisher:

Roman Kuhar

dekan Filozofske fakultete / *Dean of the Faculty of Arts*

Ljubljana, 2018

First edition

Spletno mesto konference / Conference website:

<http://www.sdjt.si/jtdh-2018/>

Publikacija je brezplačno dostopna na: / *Publication is available free of charge at:*

<http://nl.ijs.si/jtdh18/proceedings-sl.html> / <http://nl.ijs.si/jtdh18/proceedings-en.html>



To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva 4.0 Mednarodna.

This work is licensed under a Creative Commons Attribution 4.0 International License.

CIP - Kataložni zapis o publikaciji

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani

COBISS.SI-ID=296526848

ISBN 978-961-06-0111-1 (pdf)

Predgovor k zborniku konference “Jezikovne tehnologije in digitalna humanistika”

Z letošnjo konferenco obeležujemo 20 let od prve konference »Jezikovne tehnologije«, ki so jo 1998 v Cankarjevem domu v Ljubljani organizirali Tomaž Erjavec, Vojko Gorjanc, Jerneja Žganec Gros in Anica Rant. Takratna konferenca je bila posvečena uporabi in razvoju jezikovnih tehnologij za slovenščino, možnostim njihove uporabe in smernicam prihodnjega razvoja na tem področju. Na konferenci je bilo predstavljenih 26 prispevkov, ki so obravnavali govorne tehnologije in glasoslovje, uporabo računalnikov v prevajanju in poučevanju jezikov, računalniške korpuse, standarde zapisa jezikovnih podatkov in iskanje informacij na internetu. Konferenci je sledila okrogla miza, katere neposredni rezultat je bil ustanovitev Slovenskega društva za jezikovne tehnologije, ki je bil nato glavni pobudnik in organizator vseh nadaljnjih edicij konference. Skupaj s Centrom za jezikovne vire in tehnologije Univerze v Ljubljani (CJVT), Fakulteto za elektrotehniko Univerze v Ljubljani ter raziskovalnima infrastrukturama CLARIN.SI in DARIAH-SI jo 20. 9. in 21. 9. 2018 na Fakulteti za elektrotehniko prireja tudi letos, že enajstič po vrsti, ki po predlanski uspešni programski širitvi na digitalno humanistiko ohranja povezovalni fokus med disciplinama, hkrati pa si prizadeva postati pomembno srečevališče raziskovalcev v regiji.

Na letošnji konferenci se tako predstavlja 47 prispevkov, od tega 2 prispevka vabljenih predavateljev, 36 rednih polnih prispevkov in 5 povzetkov ter 4 študentski prispevki. Vse redne in študentske prispevke so pregledali trije recenzenti. 21 prispevkov je napisanih v slovenskem, 26 pa v angleškem jeziku. Skupno število vseh avtorjev prispevkov je 92. Dobra polovica avtorjev prihaja iz Slovenije, iz sosednje Hrvaške 10 %, ostali avtorji pa iz kar 19 različnih držav. Zato smo program sestavili tako, da prvi dan konference poteka v angleškem jeziku, drugi dan pa v slovenščini. Za razliko od prejšnje konference letošnje sekcije niso paralelne, kar udeležencem omogoča spremljanje vseh predavanj in predstavitev. Za ta ukrep smo se odločili prav zaradi želje po še tesnejšem sodelovanju med raziskovalci s področja jezikoslovnih tehnologij in digitalne humanistike. Istočasno smo prvič uvedli predstavitev posterjev, na kateri se jih tokrat predstavlja 9.

Urednika se iz srca zahvaljujeva vsem, ki so prispevali k uspehu konference: vabljenim predavateljem, avtorjem prispevkov, programskemu odboru za predano in natančno recenzentsko delo, organizacijskemu odboru za izvedbo konference, vodjem sekcij, da so predavanja gladko potekala, tehničnim urednikom za pripravo spletnega zbornika in sponzorjem za izkazano podporo.

Ljubljana, september 2018

Darja Fišer, Andrej Pančur

Preface to the Proceedings of the Conference on Language Technologies and Digital Humanities

With this year's conference we are celebrating the 20th anniversary since the first conference »Language technologies« which took place in 1998 in Cankarjev dom, Ljubljana and was organized by Tomaž Erjavec, Vojko Gorjanc, Jerneja Žganec Gros and Anica Rant. The topics of the first conference were the development and application of language technologies for Slovene and directions for the future. 26 papers were presented, dealing with speech technologies and phonology, computer-assisted translation and teaching, corpora, encoding standards for language data and searching for information on the internet. Following the conference a round table discussion was held, the direct result of which was the establishment of the Slovenian language technologies society which has since been the main initiator and organizer of all the following editions of the conference. Together with the Centre of language resources and technologies of the University of Ljubljana (CJVT), Faculty of Electrical Engineering of the University in Ljubljana and research infrastructures CLARIN.SI and DARIAH-SI the Society is also organizing this year's conference, held on 20-21 September 2018 at the Faculty of Electrical Engineering. In its 11th installment and after a successful expansion of the conference programme to Digital Humanities in 2016, we have retained the focus on the integration of the two disciplines and at the same time aimed to position the conference as an important meeting hub for fellow researchers in the region.

This year, 47 papers will be presented, including 2 talks by invited lecturers, 36 regular full papers and 5 abstracts, and 4 student papers. All the papers were reviewed by 3 reviewers. 21 papers were submitted in Slovene and 26 in English. The total number of all authors of the accepted papers is 92. Over half of the authors of the accepted papers are Slovene, 10% are from Croatia and the rest of the authors come from as many as 19 different countries. This is why the conference programme was designed in such a way that the first day is international, with the talks in English while talks on the second day will be held in Slovene. As opposed to the previous edition of the conference we have opted for a single track programme so that all the participants can attend all the talks, aiming to promote and foster closer collaboration among the researchers in language technologies and digital humanities. In addition, we have also introduced a poster session with 9 posters.

The editors would like to thank everyone who has contributed to the success of this conference, especially the invited lecturers and the authors of the papers for co-creating an inspiring conference programme, the Programme Committee for their dedicated reviews, the Organizing Committee for all the organizational efforts, the Session Chairs for their smooth and efficient management of the conference programme, the Technical Editors for preparing the online proceedings and the loyal sponsors for their selfless support of our activities.

Ljubljana, September 2018

Darja Fišer, Andrej Pančur

Organizacijski odbor / *Organising committee*

Simon Dobrišek, predsednik / *Chair*

Fakulteta za elektrotehniko, Univerza v Ljubljani / *Faculty of Electrical Engineering, University of Ljubljana*

Vitimir Štruc (FE)

Fakulteta za elektrotehniko, Univerza v Ljubljani / *Faculty of Electrical Engineering, University of Ljubljana*

Mojca Šorn

Inštitut za novejšo zgodovino / *Institute of Contemporary History*

Katja Zupan

Institut "Jožef Stefan" / *"Jožef Stefan" Institute*

Jerneja Fridl

Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti / *Research Centre of the Slovenian Academy of Sciences and Arts*

Programski odbor / *Programme committee*

Predsedstvo programskega odbora / *Steering committee*

Darja Fišer, predsednica / *Chair*

Filozofska fakulteta, Univerza v Ljubljani / *Faculty of Arts, University of Ljubljana*
Institut "Jožef Stefan" / *"Jožef Stefan" Institute*

Andrej Pančur

Inštitut za novejšo zgodovino / *Institute of Contemporary History*

Tomaž Erjavec

Institut "Jožef Stefan" / *"Jožef Stefan" Institute*

Simon Dobrišek

Fakulteta za elektrotehniko, Univerza v Ljubljani / *Faculty of Electrical Engineering, University of Ljubljana*

Iza Škrjanec

Študentska sekcija / *Student section*

Člani programskega odbora in recenzenti / *Programme committee members and reviewers*

- Špela Arhar Holdt**, Center za jezikovne vire in tehnologije, Univerza v Ljubljani / *Centre for language resources and technologies, University of Ljubljana*
- Zoran Bosnić**, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani / *Faculty of Computer Information Science, University of Ljubljana*
- Narvika Bovcon**, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani / *Faculty of Computer Information Science, University of Ljubljana*
- Václav Cvrček**, Inštitut češkega narodnega korpusa, Karlova univerza v Pragi / *Institute of the Czech National Corpus, Charles University in Prague*
- Helena Dobrovoljc**, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU / *Fran Ramovš Institute of the Slovenian Language, ZRC SAZU*
- Jerneja Fridl**, Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti / *Research Centre of the Slovenian Academy of Sciences and Arts*
- Polona Gantar**, Filozofska fakulteta, Univerza v Ljubljani / *Faculty of Arts, University of Ljubljana*
- Vojko Gorjanc**, Filozofska fakulteta, Univerza v Ljubljani / *Faculty of Arts, University of Ljubljana*
- Jurij Hadalin**, Inštitut za novejšo zgodovino / *Institute of Contemporary History*
- Miran Hladnik**, Filozofska fakulteta, Univerza v Ljubljani / *Faculty of Arts, University of Ljubljana*
- Ivo Ipšić**, Tehniška fakulteta, Univerza na Reki / *Faculty of Engineering, University of Rijeka*
- Mateja Jemec Tomazin**, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU / *Fran Ramovš Institute of the Slovenian Language, ZRC SAZU*
- Zdravko Kačič**, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru / *Faculty of Electrical Engineering and Computer Science, University of Maribor*
- Koraljka Kuzman Šlogar**, Inštitut za etnologijo in folkloristiko, Hrvaška / *Institute of Ethnology and Folklore Research, Croatia*
- Iztok Kosem**, Filozofska fakulteta, Univerza v Ljubljani in Institut "Jožef Stefan" / *Faculty of Arts, University of Ljubljana and "Jožef Stefan" Institute*
- Mojca Kotar**, Univerza v Ljubljani / *University of Ljubljana*
- Simon Krek**, Institut "Jožef Stefan" in Center za jezikovne vire in tehnologije, Univerza v Ljubljani / *"Jožef Stefan" Institute and Centre for Language Resources and Technologies, University of Ljubljana*
- Cvetana Krstev**, Filozofska fakulteta, Univerza v Beogradu / *Faculty of Philology, University of Belgrade*
- Drago Kunej**, Glasbenonarodopisni inštitut, ZRC SAZU / *Institute of Ethnomusicology, ZRC SAZU*
- Nikola Ljubešić**, Institut "Jožef Stefan" / *Jožef Stefan Institute*
- Nataša Logar**, Fakulteta za družbene vede, Univerza v Ljubljani / *Faculty of Social Sciences, University of Ljubljana*
- Matija Marolt**, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani / *Faculty of Computer Information Science, University of Ljubljana*
- France Mihelič**, Fakulteta za elektrotehniko, Univerza v Ljubljani / *Faculty of Electrical Engineering, University of Ljubljana*

Maja Miličević, Filološka fakulteta, Univerza v Beogradu / *Filološka fakulteta, Univerza v Beogradu*

Dunja Mladenić, Laboratorij za umetno inteligenco, Institut "Jožef Stefan" / *Artificial Intelligence Laboratory, Institute "Jožef Stefan"*

Matija Ogrin, Inštitut za slovensko literaturo in literarne vede ZRC SAZU / *Institute of Slovene Literature and Literary Sciences, ZRC SAZU*

Karmen Pižorn, Pedagoška Fakulteta, Univerza v Ljubljani / *Faculty of Education, University of Ljubljana*

Dan Podjed, Inštitut za slovensko narodopisje ZRC SAZU / *Institute of Slovenian Ethnology, ZRC SAZU*

Marko Robnik-Šikonja, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani / *Faculty of Computer Information Science, University of Ljubljana*

Tanja Samardžić, Univerza v Zurichu / *University of Zurich*

Miha Seručnik, Zgodovinski inštitut Milka Kosa ZRC SAZU / *Milko Kos Historical Institute, ZRC SAZU*

Franc Solina, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani / *Faculty of Computer Information Science, University of Ljubljana*

Marko Stabej, Filozofska fakulteta, Univerza v Ljubljani / *Faculty of Arts, University of Ljubljana*

Kristina Štrkalj Despot, Inštitut za hrvaški jezik in jezikoslovje, Hrvaška / *Institute of Croatian Language and Linguistics, Croatia*

Mojca Šorn, Inštitut za novejšo zgodovino / *Institute of Contemporary History*

Jan Šnajder, Fakulteta za elektrotehniko in računalništvo, Univerza v Zagrebu / *Faculty of Electrical Engineering and Computing, University of Zagreb*

Janez Štebe, Fakulteta za družbene vede, Univerza v Ljubljani / *Faculty of Social Sciences, University of Ljubljana*

Simon Šuster, Raziskovalni center za računalniško jezikoslovje in psiholingvistiko, Univerza v Antwerpnu / *Computational Linguistics & Psycholinguistics Research Center, University of Antwerp*

Toma Tasovac, Beograjski center za digitalno humanistiko in Trinity College Dublin / *Belgrade Center for Digital Humanities and Trinity College Dublin*

Darinka Verdonik, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru / *Faculty of Electrical Engineering and Computer Science, University of Maribor*

Aleš Vaupotič, Raziskovalni center za humanistiko, Univerza v Novi Gorici / *Research Centre for Humanities, University of Nova Gorica*

Špela Vintar, Filozofska fakulteta, Univerza v Ljubljani / *Faculty of Arts, University of Ljubljana*

Jerneja Žganec Gros, Alpineon d.o.o. / *Alpineon Ltd., Slovenia*

Organizatorji / Organizers

SDJT 

cjvt

CLARIN.SI 

 **DARIAH-SI**



Univerza v Ljubljani
Fakulteta *za elektrotehniko*

Sponzorji / Sponsors

alpineon))



UNIPOINT - DR

mikro**grafija**

Urnik / Timetable

Četrtek, 20. 9. 2018 / Thursday 20-9-2018

08.00-08.30	Registracija / Registration
08.30-09.00	Otvoritev / Opening
09.00-10.00	Vabljeni predavanja / Invited lecture Malvina Nissim, <i>Too good to be true: Current approaches to author profiling</i>
10.00-10.30	Sekcija 1 / Session 1: Strojno prevajanje / Machine translation Gregor Donaj and Mirjam S. Maučec: <i>From statistical machine translation to translation with neural networks for the Slovene-English language pair</i> Mihael Arčan: <i>A comparison of Statistical and Neural Machine Translation for Slovene, Serbian and Croatian</i>
10.30-11.00	Odmor za kavo / Coffee break
11.00-12.30	Sekcija 2 / Session 2: Jezikovni viri / Language resources Darinka Verdonik: <i>Corpus and database GOS Videlectures</i> Maria Jose Bocorny Finatto, Paulo Quaresma and Maria Filomena Gonçalves: <i>Portuguese Corpora of the 18th century: old Medicine texts for teaching and research activities</i> Nikola Ljubešić, Željko Agić, Filip Klubička, Vuk Batanović and Tomaž Erjavec: <i>hr500k -- A Reference Training Corpus of Croatian</i> Nikola Ljubešić, Darja Fišer, Tomaž Erjavec and Filip Dobranić: <i>The Parlameter corpus of contemporary Slovene parliamentary proceedings</i> Kaja Dobrovoljc: <i>N-gram Frequency Lists for Reference Corpora of Slovenian Language</i> Polona Gantar, Kristina Štrkalj Despot, Simon Krek and Nikola Ljubešić: <i>Towards Semantic Role Labeling in Slovene and Croatian</i>
12.30-13.30	Odmor za kosilo / Lunch break
13.30-14.30	Posterji + odmor za kavo / Poster session + coffee break Ajda Pretnar and Dan Podjed: <i>Data Mining Workspace Sensors: A New Approach to Anthropology</i> Petar Božović, Tomaž Erjavec, Jörg Tiedemann, Nikola Ljubešić and Vojko Gorjanc: <i>Opus-MontenegrinSubs 1.0: First electronic corpus of the Montenegrin language</i> Sara Ries: <i>Online database in Research of Correspondence of Franjo Ksaver Kuhač (1834-1911)</i>

Eneja Osrajnik, Darja Fišer and Vojko Gorjanc:
Korpusna analiza nestandardne vejice po uvajalnih prislovnih zvezah v slovenskih formalnih in neformalnih besedilih

Christof Schöch, Maciej Eder, Carolin Odebrecht, Mike Kestemont, Antonija Primorac, Justin Tonra, Katja Mihurko Poniž and Catherine Kanellopoulou:
Distant Reading for European Literary History. A COST Action

Katja Mihurko Poniž, Amelia Sanz, Marie Nedregotten Sørbo, Suzan van Dijk, Viola Parente-Čapková, Narvika Bovcon and Aleš Vaupotič:
Teaching women writers with NEWW Virtual Research Environment

Nikola Ljubešić, Tomaž Erjavec and Darja Fišer:
KAS-term and KAS-biterm: Datasets and baselines for monolingual and bilingual terminology extraction from academic writing

Benedikt Perak and Filip Rodik:
Building a corpus of the Croatian parliamentary debates using ReLDI API for tokenization, lemmatization, syntactic parsing and Neo4j graph database for creation of social ontology model, text classification and extraction of semantic information

Damjan Popič and Darja Fišer:
Odnosi do jezika v slovenski, hrvaški in srbski računalniško posredovani komunikaciji

14.30-16.00

Sekcija 3 / Session 3: Digitalna humanistika / Digital humanities

Ivanka Rajh and Siniša Runjaić:
Crowdsourcing terminology: harnessing the potential of translator's glossaries

Dan Podjed and Ajda Pretnar:
Self-Promotion on Instagram: A Case of President's Profile

Lana Hudeček and Milica Mihaljević:
One year later - What is different?

Darja Fišer and Monika Kalin Golob:
A corpus analysis of tweets of Slovene corporate users

Tobias Weber and Jeremy Bradley:
Exploring Finno-Ugric linguistics through solving IT problems

Narvika Bovcon and Aleš Vaupotič:
Artistic Visualizations and Beyond: A Study of Materializations of a Digital Database

16.00-17.30

Sekcija 4 / Session 4: Jezikovne tehnologije / Language technologies

Milan van Lange and Ralf Futselaar:
Debating Evil: Using Word Embeddings to Analyze Parliamentary Debates on War Criminals in The Netherlands

Tatjana Marvin, Jure Derganc, Samo Beguš and Saba Battelino:
Word Selection in the Slovenian Sentence Matrix Test for Speech Audiometry

Aleksander Ključevšek, Simon Krek and Marko Robnik-Šikonja:
Efficient calculation of frequency statistics for Slovene language corpora

Vuk Batanović, Nikola Ljubešić and Tanja Samardžić:
SETimes.SR - A Reference Training Corpus of Serbian

Tadej Škvorc, Simon Krek, Senja Pollak, Spela Arhar Holdt and Marko Robnik-Šikonja:
Evaluation of Statistical Readability Measures on Slovene texts

Aniko Kovač and Maja Markovič:
A Rule-Based Syllabifier for Serbian

17.30-19.00 **Občni zbor + kava / General assembly + coffee**

20.00-22.00 **Večerja / Dinner**

Petek, 21. 9. 2018 / Friday 21-9-2018

08.30-09.00 **Registracija / Registration**

09.00-10.00 **Vabljeni predavanja / Invited lecture**
Martijn Kleppe,
Bringing Digital Humanities to the wider public: libraries as incubator for DH research results

10.00-10.30 **Sekcija 5 / Session 5: Infrastruktura / Infrastructure**
Maja Dolinar, Janez Štebe and Sonja Bezjak:
Razvoj smernic za predajo in arhiviranje kvalitativnih podatkov v Arhivu družboslovnih podatkov

Darja Fišer, Jakob Lenardič and Tomaž Erjavec:
Citiranje jezikoslovnih podatkov v slovenskih znanstvenih objavah: stanje in priporočila

10.30-11.00 **Odmor za kavo / Coffee break**

11.00-12.30 **Sekcija 6 / Session 6: Korpusno jezikoslovje / Corpus linguistics**
Nataša Logar and Tomaž Erjavec:
Strokovno-znanstvena slovenščina: besednovrstne in oblikoskladenjske značilnosti

Peter Holozan:
Zbirka primerov rabe vejice Vejica 1.3

Miha Pavlovič and Rena Ito:
Korpusna analiza slovničnih napak v spisih učencev japonščine na osnovni ravni

Helena Dobrovoljc and Urška Vranjek Ošlak:
Zakaj ne z eno poizvedbo po različnih korpusih? (Troje korpusnih preverb pod primerjalnim drobnogledom)

Iztok Kosem, Simon Krek, Polona Gantar, Špela Arhar Holdt, Jaka Čibej and Cyprian Laskowski:
Kolokacijski slovar sodobne slovenščine

Polona Gantar, Špela Arhar Holdt, Jaka Čibej, Taja Kuzman and Teja Kavčič:
Glagolske večbesedne enote v učnem korpusu ssj500k 2.1

12.30-13.30	Odmor za kosilo / Lunch break
13.30-14.30	Študentska sekcija + odmor za kavo / Student session + Coffee break Urška Bratoš: <i>Gradnja korpusa tvitov slovenskih politikov Janes-TwePo</i> Isolde Van Dorst: <i>You, thou and thee: A statistical analysis of Shakespeare's use of pronominal address terms</i> Gabi Rolih: <i>K-means Clustering of CMC Data for Tagger Improvement</i> Klara Eva Kukovičič: <i>Uporabnost luščilnikov terminologije Sketch Engine in CollTerm z vidika (študenta) prevajalca</i>
14.30-15.30	Sekcija 7 / Session 7: Digitalna humanistika / Digital humanities Andrej Pančur: <i>Trajnost digitalnih izdaj: Uporaba statičnih spletnih strani na portalu Zgodovina Slovenije - Sistory</i> Andrej Pančur, Alenka Pirman and Maruša Kocjančič: <i>Spregledana kulturna dediščina in uporaba digitalne raziskovalne infrastrukture za humanistiko v raziskavi Odlivanje smrti</i> Alenka Kavčič, Ivan Lovrić and Vera Smole: <i>Interaktivna karta slovenskih narečnih besedil</i> Nina Ditmajer, Matija Ogrin and Tomaž Erjavec: <i>Zapis in prikaz starejših pesniških besedil ter njihovih variant v TEI</i>
15.30-16.00	Zaključek / Closing

Kazalo / Table of Contents

Predgovor	i
<i>Preface</i>	ii
Organizacijski odbor / Organising committee	iii
Programski odbor / Programme committee	iii
Predsedstvo programskega odbora / Steering committee	iii
Člani programskega odbora in recenzenti / Programme committee members and reviewers	iv
Organizatorji / Organizers	vi
Sponsorji / Sponsors	vii
Urniki / Timetable	viii
Kazalo / Table of Contents	xii
VABLJENI PRISPEVKI / INVITED TALKS	1
Too good to be true: Current approaches to author profiling <i>Malvina Nissim</i>	1
Bringing Digital Humanities to the wider public: libraries as incubator for DH research results <i>Martijn Kleppe</i>	2
PRISPEVKI – PAPERS	3
A comparison of Statistical and Neural Machine Translation for Slovene, Serbian and Croatian <i>Mihael Arčan</i>	3
SETimes.SR - A Reference Training Corpus of Serbian <i>Vuk Batanović, Nikola Ljubešić and Tanja Samardžić</i>	11
Artistic Visualizations and Beyond: A Study of Materializations of a Digital Database <i>Narvika Bovcon and Aleš Vaupotič</i>	18
Opus-MontenegrinSubs 1.0: First electronic corpus of the Montenegrin language <i>Petar Božović, Tomaž Erjavec, Jörg Tiedemann, Nikola Ljubešić and Vojko Gorjanc</i>	24
Zapis in prikaz starejših pesniških besedil ter njihovih variant v TEI <i>Nina Ditmajer, Matija Ogrin and Tomaž Erjavec</i>	29
Zakaj ne z eno poizvedbo po različnih korpusih? (Troje korpusnih preverb pod primerjalnim drobnogledom) <i>Helena Dobrovoljc and Urška Vranjek Ošlak</i>	38

Frekvenčni sezname n-gramov v korpusih slovenskega jezika	
<i>Kaja Dobrovoljc</i>	47
Razvoj smernic za predajo in arhiviranje kvalitativnih podatkov v Arhivu družboslovnih podatkov	
<i>Maja Dolinar, Janez Štebe and Sonja Bezjak</i>	55
Prehod iz statističnega strojnega prevajanja na prevajanje z nevronskimi omrežji za jezikovni par slovenščina-angleščina	
<i>Gregor Donaj and Mirjam S. Maučec</i>	62
Analiza tvitov slovenskih korporativnih uporabnikov	
<i>Darja Fišer and Monika Kalin Golob</i>	69
Citiranje jezikoslovnih podatkov v slovenskih znanstvenih objavah: stanje in priporočila	
<i>Darja Fišer, Jakob Lenardič and Tomaž Erjavec</i>	77
Glagolske večbesedne enote v učnem korpusu ssj500k 2.1	
<i>Polona Gantar, Špela Arhar Holdt, Jaka Čibej, Taja Kuzman and Teja Kavčič</i>	85
Towards Semantic Role Labeling in Slovene and Croatian	
<i>Polona Gantar, Kristina Štrkalj Despot, Simon Krek and Nikola Ljubešić</i>	93
Zbirka primerov rabe vejice Vejica 1.3	
<i>Peter Holozan</i>	99
Croatian Web Dictionary Mrežnik: One year later - What is different?	
<i>Lana Hudeček and Milica Mihaljević</i>	106
Portuguese Corpora of the 18th century: old Medicine texts for teaching and research activities	
<i>Maria Jose Bocorny Finatto, Paulo Quaresma and Maria Filomena Gonçalves</i>	114
Interaktivna karta slovenskih narečnih besedil	
<i>Alenka Kavčič, Ivan Lovrić and Vera Smole</i>	121
Učinkovit izračun frekvenčnih statistik za slovenske jezikovne korpuse	
<i>Aleksander Ključevšek, Simon Krek and Marko Robnik-Šikonja</i>	126
Kolokacijski slovar sodobne slovenščine	
<i>Iztok Kosem, Simon Krek, Polona Gantar, Špela Arhar Holdt, Jaka Čibej and Cyprian Laskowski</i>	133
A Rule-Based Syllabifier for Serbian	
<i>Aniko Kovač and Maja Marković</i>	140
Debating Evil: Using Word Embeddings to Analyze Parliamentary Debates on War Criminals in The Netherlands	
<i>Milan van Lange and Ralf Futselaar</i>	147

hr500k – A Reference Training Corpus of Croatian	
<i>Nikola Ljubešić, Željko Agić, Filip Klubička, Vuk Batanović and Tomaž Erjavec</i>	154
The Parlameter corpus of contemporary Slovene parliamentary proceedings	
<i>Nikola Ljubešić, Darja Fišer, Tomaž Erjavec and Filip Dobranić</i>	162
KAS-term and KAS-biterm: Datasets and baselines for monolingual and bilingual terminology extraction from academic writing	
<i>Nikola Ljubešić, Tomaž Erjavec and Darja Fišer</i>	168
Strokovno-znanstvena slovenščina: besednovrstne in oblikoskladenjske značilnosti	
<i>Nataša Logar and Tomaž Erjavec</i>	175
Word Selection in the Slovenian Sentence Matrix Test for Speech Audiometry	
<i>Tatjana Marvin, Jure Derganc, Samo Beguš and Saba Battelino</i>	181
Korpusna analiza nestandardne vejice po uvajalnih prislovnih zvezah v slovenskih formalnih in neformalnih besedilih	
<i>Eneja Osrajnik, Darja Fišer and Vojko Gorjanc</i>	188
Trajnost digitalnih izdaj: Uporaba statičnih spletnih strani na portalu Zgodovina Slovenije - Sistory	
<i>Andrej Pančur</i>	196
Spre gledana kulturna dediščina in uporaba digitalne raziskovalne infrastrukture za humanistiko v raziskavi Odlivanje smrti	
<i>Andrej Pančur, Alenka Pirman and Maruša Kocjančič</i>	203
Analiza slovničnih napak v korpusu spisov učencev japonščine na osnovni ravni	
<i>Miha Pavlovič and Rena Ito</i>	211
Building a corpus of the Croatian parliamentary debates using ReLDI API for tokenization, lemmatization, syntactic parsing and Neo4j graph database for creation of social ontology model, text classification and extraction of semantic information	
<i>Benedikt Perak and Filip Rodik</i>	216
Samopromocija na Instagramu: Primer predsednikovega profila	
<i>Dan Podjed and Ajda Pretnar</i>	221
Data Mining Workspace Sensors: A New Approach to Anthropology	
<i>Ajda Pretnar and Dan Podjed</i>	227
Crowdsourcing terminology: harnessing the potential of translator’s glossaries	
<i>Ivanka Rajh and Siniša Runjaić</i>	234
Evaluation of Statistical Readability Measures on Slovene texts	
<i>Tadej Škvorc, Simon Krek, Senja Pollak, Spela Arhar Holdt and Marko Robnik-Šikonja</i>	240
Exploring Finno-Ugric linguistics through solving IT problems	
<i>Tobias Weber and Jeremy Bradley</i>	248

POVZETKI – ABSTRACTS	254
Teaching women writers with NEWW Virtual Research Environment	
<i>Katja Mihurko Poniž, Amelia Sanz, Marie Nedregotten Sørbo, Suzan van Dijk, Viola Parente-Čapková, Narvika Bovcon and Aleš Vaupotič</i>	254
Odnosi do jezika v slovenski, hrvaški in srbski računalniško posredovani komunikaciji	
<i>Damjan Popič and Darja Fišer</i>	256
Online database in Research of Correspondence of Franjo Ksaver Kuhač (1834-1911)	
<i>Sara Ries</i>	260
Distant Reading for European Literary History. A COST Action	
<i>Christof Schöch, Maciej Eder, Carolin Odebrecht, Mike Kestemont, Antonija Primorac, Justin Tonra, Katja Mihurko Poniž and Catherine Kanellopoulou</i>	262
Corpus and database GOS Videolectures	
<i>Darinka Verdonik</i>	265
ŠTUDENSKI PRISPEVKI – STUDENT PAPERS	269
Gradnja korpusa tvitov slovenskih politikov Janes-TwePo	
<i>Urška Bratoš</i>	269
You, thou and thee: A statistical analysis of Shakespeare's use of pronominal address terms	
<i>Isolde Van Dorst</i>	274
Uporabnost luščilnikov terminologije Sketch Engine in CollTermz vidika (študenta) prevajalca	
<i>Klara Eva Kukovičič</i>	281
ŠTUDENSKI POVZETKI – STUDENT ABSTRACTS	288
K-means Clustering of CMC Data for Tagger Improvement	
<i>Gabi Rolih</i>	288