

Zbornik konference
Jezikovne tehnologije in digitalna humanistika

Proceedings "qh"vj g'Eqphgt gpeg"qp
Language Technologies & Digital Humanities

29. september – 1. oktober 2016
Filozofska fakulteta, Univerza v Ljubljani
Ljubljana, Slovenija

September 29th – October 1st, 2016
Faculty of Arts, University of Ljubljana
Ljubljana, Slovenia

Uredila / *Edited by*
Tomaž Erjavec, Darja Fišer

ZBORNİK KONFERENCE
JEZIKOVNE TEHNOLOGIJE IN DIGITALNA HUMANISTIKA

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

Uredila / Edited by: Tomaž Erjavec, Darja Fišer
Tehnični uredniki / Technical editors: Jaka Čibej, Katja Zupan

Založil / Published by:
*Znanstvena založba Filozofske fakultete v Ljubljani /
Ljubljana University Press, Faculty of Arts*

Izdal / Issued by:
Institut »Jožef Stefan«, Ljubljana / *Jožef Stefan Institute, Ljubljana*
Oddelek za prevajalstvo / *Department of Translation Studies*

Za založbo / For the publisher:
Branka Kalenić Ramšak,
dekanja Filozofske fakultete / *Dean of the Faculty of Arts*

Ljubljana, 2016
Prva izdaja / *First edition*

Spletno mesto konference / *Conference web site:* <http://www.sdjt.si/jtdh-2016/>

Publikacija je brezplačno dostopna na: / *Publication is available free of charge at:*
<http://nl.ijs.si/isjt16/proceedings-sl.html> / <http://nl.ijs.si/isjt16/proceedings-en.html>



To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna.

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International license.

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=286548224
ISBN 978-961-237-862-2 (pdf)

Predgovor k zborniku konference “Jezikovne tehnologije in digitalna humanistika”

Prva konferenca Jezikovne tehnologije je potekala davnega leta 1998 in se nato redno nadaljevala vsaki dve leti. Ob jubilejni deseti konferenci JT v letu 2016 smo se odločili za programsko širitev na področje digitalne humanistike, ki je kot presek digitalnih tehnologij in humanistike aktualno raziskovalno področje, kjer se digitalne tehnologije na eni strani uporabljajo pri raziskavah v humanistiki za študij jezika, družbe in kulture, na drugi strani pa humanistika spodbuja razvoj novih tehnoloških rešitev. Digitalna humanistika je mednarodno že uveljavljeno področje, kot npr. kažejo revija *Journal of Digital Humanities*, zveza *Alliance of Digital Humanities Organizations* s svojimi letnimi konferencami *Digital Humanities in* mreža *NeDiMAH* (*Network for Digital Methods in the Arts and Humanities*). V samem izhodišču je digitalna humanistika izrazito interdisciplinarno in kolaborativno raziskovalno delo, ki bistveno spreminja ustaljene humanistične pristope ter spodbuja razvoj novih analitičnih tehnik in metod, v slovenskem prostoru pa še nima skupnega mesta za predstavitev dosežkov svojega dela in diskusijo med različnimi deležniki na področju. Zato so Slovensko društvo za jezikovne tehnologije (SDJT), Center za jezikovne vire in tehnologije Univerze v Ljubljani (CJVT) ter raziskovalni infrastrukturi CLARIN.SI in DARIAH-SI 29. 9.–1. 10. 2016 na Filozofski fakulteti Univerze v Ljubljani organizirale konferenco Jezikovne tehnologije in digitalna humanistika.

Za razliko od konferenc JT, in glede na specifikke DH, tokrat prispevkov nismo omejili na polne članke, temveč smo sprejemali tudi razširjene povzetke in uvedli študentsko sekcijo, v spremljevalni konferenčni program pa uvrstili še panel Terminologija v poklicnem vsakdanu: stanje in potrebe in sekcijo s predstavitvijo sponzorjev.

Vse redne in študentske prispevke sta pregledala dva recenzenta, k sodelovanju pa nam je uspelo pritegniti tudi pet vabljenih predavateljev. Zbornik vsebuje 58 sprejetih prispevkov, od tega 5 prispevkov vabljenih predavateljev, 30 rednih polnih prispevkov in 17 povzetkov ter 6 študentskih prispevkov, pri čemer je 41 prispevkov v slovenskem, ostali pa v angleškem, hrvaškem, srbskem in bosanskem jeziku.

Urednika se zahvaljujeta vsem, ki so prispevali k uspehu konference: vabljenim predavateljem, avtorjem prispevkov, programskemu odboru za izjemno kvalitetno recenzentsko delo, organizacijskemu odboru za izvedbo konference, predsednikom sekcij, da so predavanja gladko potekala, tehničnim urednikom za pripravo spletnega zbornika in sponzorjem za izkazano podporo.

Ljubljana, september 2016

Tomaž Erjavec, Darja Fišer

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

The first Language Technologies Conference in Slovenia took place in the long-past year 1998 and from then on regularly every second year. On the occasion of the anniversary tenth conference in 2016 we decided to widen its scope to the field of Digital Humanities. DH is, as the intersection of digital technologies and the humanities, a highly relevant research field, where, on the one hand, digital technologies are used in the study of language, society and culture, and, on the other hand, humanities research paves the way for the development of new digital technologies. Digital Humanities is internationally already an established field, as is evidenced by the Journal of Digital Humanities, the Alliance of Digital Humanities Organizations with its annual conferences Digital Humanities and the Network for Digital Methods in the Arts and Humanities (NeDiMAH). DH is, by definition, highly interdisciplinary and collaborative, radically changing the accepted practices in the humanities research and encouraging the development of new analytical techniques and methods but has so far lacked a national or regional event to present its results and encourage discussion.

To this end, the Slovenian Language Technologies Society (SDJT), the Centre for Language Resources and Technologies at the University of Ljubljana (CJVT), and the Slovenian research infrastructures CLARIN.SI and DARIAH-SI organised the conference Language Technologies and Digital Humanities which took place September 29th to October 1st 2016 at the Arts Faculty of the University of Ljubljana.

In contrast to the previous editions of the Language Technologies conferences, and taking into account the specifics of the Digital Humanities field, the 2016 submission were not limited to full papers, but extended abstracts were accepted as well. Additionally, a student session, a round table on terminology and a session with the presentation of the sponsors was organised. All regular and student contributions were reviewed by two members of the programme committee, with the conference also presenting the talks of five invited speakers. The proceedings contain 58 accepted papers, 5 of which are from invited speakers, 30 regular full papers, 17 abstracts, and 6 student contributions, with 41 contributions in Slovenian, with the others in English, Croatian, Serbian, and Bosnian languages.

The editors would like to thank all the people who contributed to the success of the conference: the invited speakers, authors of the papers and abstracts, the programme committee for exemplary reviewing of the contributions, the organising committee for the smooth functioning of the event, sessions chairs who made sure the talks proceeded smoothly, the technical editors for the preparation of the proceedings and the sponsors for supporting the event.

Ljubljana, September 2016

Tomaž Erjavec, Darja Fišer

Organizacijski odbor / Organising committee

Darja Fišer, predsednica / Chair (SDJT, FF, IJS)

Jerneja Fridl (DARIAH-SI, ZRC SAZU)

Vojko Gorjanc (CJVT, FF)

Simon Krek (CJVT, CLARIN.SI, IJS)

Mojca Šorn (DARIAH-SI, INZ)

Kaja Dobrovoljc (študentska sekcija / student session, SDJT)

Programski odbor / Programme committee

Predsedstvo programskega odbora / Steering committee

Tomaž Erjavec, predsednik / Chair

Institut "Jožef Stefan" / Jožef Stefan Institute (CLARIN.SI)

Darja Fišer

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

Vojko Gorjanc

Filozofska fakulteta, Univerza v Ljubljani / Faculty of Arts, University of Ljubljana (CJVT)

Andrej Pančur

Inštitut za novejšo zgodovino / Institute of Contemporary History (DARIAH-SI)

Vodje tematskih področij / Area chairs

Govorne tehnologije / Speech technologies:

Jerneja Žganec Gros, Alpineon d.o.o.

Jezikovne tehnologije / Language technologies:

Tomaž Erjavec, Institut "Jožef Stefan" / "Jožef Stefan" Institute

Prevodoslovje / Translation studies:

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

Leksikologija in leksikografija / Lexicology and lexicography:

Simon Krek, Institut "Jožef Stefan" / Center za jezikovne vire in tehnologije, Univerza v Ljubljani / Jožef Stefan Institute / Center for Language Resources and Technologies, University of Ljubljana

Jezikovna standardizacija / Language standardisation:

Helena Dobrovoljc, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU / Fran Ramovš Institute of Slovenian Language, Scientific Research Centre of the Slovenian Academy of Sciences and Arts (ZRC SAZU)

Jezikovna didaktika in opismenjevanje / Didactics and literacy pedagogy:

Špela Arhar Holdt, Trojina, zavod za uporabno slovenistiko / Trojina, Institute for Applied Slovene Studies and Filozofska fakulteta, Univerza v Ljubljani / Faculty of Arts, University of Ljubljana

Izobraževanje / Education:

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

Literarne vede / Literary studies:

Miran Hladnik, Filozofska fakulteta, Univerza v Ljubljani / Faculty of Arts, University of Ljubljana

Zgodovinopisje / History studies:

Andrej Pančur, Inštitut za novejšo zgodovino / Institute of Contemporary History

Tehnološke rešitve za raziskave in predstavitve kulturne dediščine / Technology solutions for research and presentation of cultural heritage:

Jerneja Fridl, ZRC SAZU / Scientific Research Centre of the Slovenian Academy of Sciences and Arts

Digitalne umetnosti in prakse / Digital arts and practices:

Jurij Hadalin, Inštitut za novejšo zgodovino / Institute of Contemporary History

Diseminacija in arhiviranje digitalnih vsebin / Dissemination and archiving of digital resources:

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

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

Zoran Bosnić, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani / Faculty of Computer and Information Science, University of Ljubljana

Narvika Bovcon, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani / Faculty of Computer and Information Science, University of Ljubljana

Václav Cvrček, Inštitut češkega narodnega korpusa, Karlova univerza v Pragi / Czech National Corpus Institute, Charles University in Prague

Vlado Delić, Fakulteta za tehnične vede, Univerza v Novem Sadu / Faculty of Technical Sciences, University of Novi Sad

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

Polona Gantar, Filozofska fakulteta, Univerza v Ljubljani / Faculty of Arts, University of Ljubljana

Tatjana Hajtnik, Arhiv Republike Slovenije / Archive of the Republic of Slovenia

Mario Hibert, Filozofska fakulteta, Univerza v Sarajevu / Faculty of Arts, University of Sarajevo

Ivo Ipšić, Tehniška fakulteta, Univerza na Reki / Technical Faculty, University of Rijeka

Mateja Jemec Tomazin, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU / Fran Ramovš Institute of 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

Iztok Kosem, Trojina, zavod za uporabno slovenistiko / Trojina, Institute for Applied Slovene Studies in/and Filozofska fakulteta, Univerza v Ljubljani / Faculty of Arts, University of Ljubljana

Žiga Kokalj, Inštitut za antropološke in prostorske študije ZRC SAZU / Institute of Anthropological and Spatial Studies, ZRC SAZU

Mojca Kotar, Univerza v Ljubljani / University of Ljubljana

Cvetana Krstev, Filozofska fakulteta, Univerza v Beogradu / Faculty of Arts, University of Beograd

Drago Kunej, Glasbenonarodopisni inštitut, ZRC SAZU / Institute of Ethnomusicology, ZRC SAZU

Nikola Ljubešić, Odsek za informacijske in komunikacijske znanosti, Univerza v Zagrebu / Department of Information and Communication Sciences, University of Zagreb

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 and 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 / Faculty of Philology, University of Belgrade

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

Matija Ogrin, Inštitut za slovensko literaturo in literarne vede ZRC SAZU / Institute of Slovenian Literature and Literary Studies, ZRC SAZU

Miha Peče, Inštitut za slovensko narodopisje ZRC SAZU / Institute of Slovenian Ethnology, ZRC SAZU

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 and 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 in/and Information Science, University of Ljubljana

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

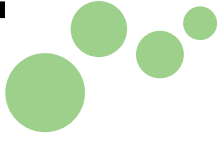
Jan Šnajder, Fakulteta za elektrotehniko in računalništvo, Univerza v Zagrebu / Faculty of Electrical Engineering and Computer Science, University of Zagreb

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

Benjamin Štular, Inštitut za arheologijo ZRC SAZU / Institute of Archaeology, ZRC SAZU

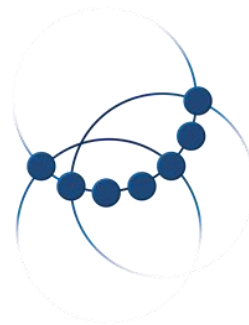
Darinka Verdonik, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru / Faculty of Electrical Engineering and Computer Science, University of Maribor

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

SDJT 

C _____ **J** _____
V _____
T _____

CLARIN.SI



 **DARIAH-SI**

Sponzorji / Sponsors

alpineon))

invita

mikro**grafija**

UNIPOINT - DR

računalniški inženiring, d.o.o.

Urnik / Timetable

Četrtek, 29. 9. 2016 / Thursday 29-9-2016

8.30-9.00	Registracija / Registration
9.00-9.30	Otvoritev / Opening (predavalnica / room: 34)
9.30-10.30	Vabljeno predavanje / Invited lecture (predavalnica / room: 34) Daniel Zeman, <i>Universal Dependencies for Slavic Languages</i>
10.30-11.00	Odmor za kavo / Coffee break
11.00-12.30	Paralelna sekcija 1A / Parallel session 1A (predavalnica / room: 34)
<i>11.00 - 11.20</i>	Mihael Arčan, Maja Popović, Paul Buitelaar: <i>Asistent – A Machine Translation System for Slovene, Serbian and Croatian</i>
<i>11.20 - 11.40</i>	Shaun Azzopardi, Albert Gatt, Gordon Pace: <i>Integrating Natural Language and Formal Analysis for Legal Documents</i>
<i>11.40 - 12.00</i>	Nikola Ljubešić, Tomaž Erjavec, Darja Fišer, Tanja Samardžić, Maja Miličević, Filip Klubička, Filip Petkovski: <i>Easily Accessible Language Technologies for Slovene, Croatian and Serbian</i>
<i>12.00 - 12.15</i>	Nikola Ljubešić, Tanja Samardžić, Maja Miličević: <i>Analysing Spatial Distribution of Linguistic Variables in Geocoded Tweets from Croatia, Bosnia, Montenegro and Serbia</i>
<i>12.15 - 12.30</i>	Marijana Janjić, Sara Librenjak, Kristina Kocijan: <i>Asian Language Teaching and Learning - the Influence of Technology on Students' Skills in SL Classroom</i>
11.00-12.30	Paralelna sekcija 1B / Parallel session 1B (predavalnica / room: 18)
<i>11.00 - 11.20</i>	Slobodan Mandić: <i>The First World War on the Web - The Case of Serbia</i>
<i>11.20 - 11.40</i>	Aleš Vaupotič, Marco Buziol, Narvika Bovcon: <i>Digital Video in Digital Humanities Methodology: A Case Study</i>
<i>11.40 - 12.00</i>	Narvika Bovcon, Jure Demšar, Aleš Vaupotič: <i>Organiziranje projekta vizualizacije podatkov</i>
<i>12.00 - 12.15</i>	Gregor Strle, Matija Marolt: <i>Language Technologies in Humanities: Computational Semantic Analysis in Folkloristics</i>
<i>12.15 - 12.30</i>	Mario Hibert: <i>What Is Critical in Digital Humanities?</i>
12.30-13.30	Odmor za kosilo / Lunch break
13.30-14.30	Vabljeno predavanje / Invited lecture (predavalnica / room: 34) Niels-Oliver Walkowski, <i>The Landscape of Digital Annotations and Its Meaning</i>
14.30-15.00	Odmor za kavo / Coffee break

15.00-16.30	Paralelna sekcija 2A / Parallel session 2A (predavalnica / room: 34)
<i>15.00 - 15.20</i>	Darja Fišer, Jasmina Smailović, Tomaž Erjavec, Miha Grčar, Igor Mozetič: <i>Sentiment Annotation of the Janes Corpus of Slovene User-Generated Content</i>
<i>15.20 - 15.40</i>	Natalia Korchagina: <i>Building a Gold Standard for Temporal Entity Extraction from Medieval German Texts</i>
<i>15.40 - 16.00</i>	Ivana Lalli Pačelat: <i>Priprava usporedivih korpusa za usporedbu</i>
<i>16.00 - 16.15</i>	Stevan Ostrogonac, Branislav Popović, Milan Sečujski: <i>The Use of Semantic Word Classes in Document Classification</i>
<i>16.15 - 16.30</i>	Tadeja Rozman, Špela Arhar Holdt, Senja Pollak, Iztok Kosem: <i>Luščenje in jezikoslovna analiza kolokacij iz korpusa Šolar</i>
15.00-16.30	Paralelna sekcija 2B / Parallel session 2B (predavalnica / room: 18)
<i>15.00 - 15.20</i>	Anja Ragolič: <i>Eagle - Medomrežje Europeana antične grške in latinske epigrafike. Digitalni dostop do antičnih napisnih spomenikov</i>
<i>15.20 - 15.40</i>	Maja Vičič Krabonja: <i>Digitalna humanistika v šoli</i>
<i>15.40 - 16.00</i>	Andrej Pančur: <i>Popisi prebivalstva Slovenije 1830–1931: Orodje za transkribiranje historičnih demografskih podatkov</i>
<i>16.00 - 16.15</i>	Aleš Lazar, Sonja Ifko: <i>Trirazsežno dokumentiranje v službi varovanja nepremične kulturne dediščine</i>
<i>16.15 - 16.30</i>	Andrej Pančur, Bogomir Rožman: <i>Dolgotrajno ohranjanje raziskovalnih podatkov v manjših raziskovalnih infrastrukturah: Uporaba odprtokodne aplikacije Archivematica</i>
16.30-17.30	Sponzorska sekcija / Presentations of sponsors (predavalnica / room: 18)
Petek, 30. 9. 2016 / Friday 30-9-2016	
9.00-9.30	Registracija / Registration
9.30-10.30	Vabljeni predavanja / Invited lecture (predavalnica / room: 34) Laurent Romary, <i>The Text Encoding Initiative: 30 Years of Accumulated Wisdom and Its Potential for a Bright Future</i>
10.30-11.00	Odmor za kavo / Coffee break
11.00-12.30	Paralelna sekcija 3A / Parallel session 3A (predavalnica / room: 34)
<i>11.00 - 11.20</i>	Jaka Čibej, Špela Arhar Holdt, Tomaž Erjavec, Darja Fišer: <i>Razvoj učne množice za izboljšano označevanje spletnih besedil</i>
<i>11.20 - 11.40</i>	Polona Gantar, Iza Škrjanec, Darja Fišer, Tomaž Erjavec: <i>Slovar tviterščine</i>

11.40 - 12.00	Špela Arhar Holdt, Kaja Dobrovoljc, Iztok Kosem: <i>Predstavitveni portal spletnih jezikovnih virov za slovenščino</i>
12.00 - 12.15	Kaja Dobrovoljc, Tomaž Erjavec, Simon Krek: <i>Pretvorba korpusa ssj500k v Univerzalno odvisnostno drevesnico za slovenščino</i>
12.15 - 12.30	Simon Krek, Polona Gantar, Špela Arhar Holdt, Vojko Gorjanc: <i>Nadgradnja korpusov Gigafida, Kres, ccGigafida in ccKres</i>
11.00-12.30	Paralelna sekcija 3B / Parallel session 3B (predavalnica / room: 18)
11.00 - 11.20	Simon Krek, Polona Gantar, Kaja Dobrovoljc, Iza Škrjanec: <i>Označevanje udeleženskih vlog v učnem korpusu za slovenščino</i>
11.20 - 11.40	Helena Dobrovoljc: <i>Povezljivost pravopisnih pravil in slovarja: sanje pravopiscev 20. stoletja</i>
11.40 - 12.00	Andrej Pančur: <i>Označevanje zbirke zapiskov sej slovenskega parlamenta s smernicami TEI</i>
12.00 - 12.15	Bojan Kastelic, Mateja Belak, Andrej Pleterski, Benjamin Štular, Miran Erič: <i>Zbiva in EWD, spletni orodji za arheološke raziskave</i>
12.15 - 12.30	-
12.30-13.30	Odmor za kosilo / Lunch break
13.30-14.30	Vabljen predavanje / Invited lecture (predavalnica / room: 34) Peter Juel Henriksen, <i>Speech is Golden - on ASR at the Service of the Danish Public Sector</i>
14.30-15.00	Odmor za kavo / Coffee break
15.00-16.30	Paralelna sekcija 4A / Parallel session 4A (predavalnica / room: 34)
15.00 - 15.20	Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Nataša Logar, Milan Ojsteršek: <i>Slovenska akademska besedila: prototipni korpus in načrt analiz</i>
15.20 - 15.40	Klemen Kadunc, Marko Robnik-Šikonja: <i>Analiza mnenj s pomočjo strojnega učenja in slovenskega leksikona sentimenta</i>
15.40 - 16.00	Katja Zupan, Tomaž Erjavec: <i>Generiranje kritičnih prepisov s strojnim prevajanjem na ravni znakov</i>
16.00 - 16.15	Dan Podjed, Saša Babič, Tatiana Bajuk Senčar, Alenka Bezjak Mlakar, Gregor Burger, Jurij Fikfak, Jože Guna, Marko Maver, Matevž Pogačnik, Emilija Stojmenova, Uroš Žolnir: <i>Razvoj aplikacije za spodbujanje trajnostne mobilnosti</i>
16.15 - 16.30	Tatjana Veljanovski, Žiga Kokalj: <i>Slikovna retrospektiva porušenega Breginja in analiza pokrajinskih sprememb</i>
15.00-16.30	Paralelna sekcija 4B / Parallel session 4B (predavalnica / room: 18)
15.00 - 15.20	Andrej Žgank, Darinka Verdonik, Mirjam Sepesy Maučec: <i>Razpoznavanje tekočega govora v slovenščini z bazo predavanj SI TEDx-UM</i>

15.20 - 15.40	Jerneja Žganec Gros, Boštjan Vesnicher, Simon Rozman, Peter Holozan, Tomaž Šef: <i>Sintetizator govora za slovenščino eBralec</i>
15.40 - 16.00	Simon Dobrišek, David Čefarin, Vitomir Štruc, France Mihelič: <i>Preizkus Googlovega govornega programskega vmesnika pri samodejnem razpoznavanju govorne slovenščine</i>
16.00 - 16.15	Miha Seručnik: <i>Historična topografija Slovenije</i>
16.15 - 16.30	Drago Kunej: <i>Digitalizacija in uporaba digitalnega etnomuzikološkega zvočnega gradiva – izkušnje Glasbenonarodopisnega inštituta ZRC SAZU</i>
16.30-18.00	Panel o terminologiji / Round table on terminology (predavalnica / room: 34)
20.00-22.00	Večerja / Dinner
Sobota, 1. 10. 2016 / Saturday 1-10-2016	
9.30-9.45	Registracija / Registration
9.45-10.30	Vabljeno predavanje / Invited lecture (predavalnica / room: 34) Oliver Čulo, <i>Contrasting Post-editing and Human Translation along the Dimension of Term Translation</i>
10.30-11.00	Odmor za kavo / Coffee break
11.00-12.30	Paralelna sekcija 5A / Parallel session 5A (predavalnica / room: 34)
11.00 - 11.20	Benjamin Štular, Franco Niccolucci, Julian Richards: <i>ARIADNE: povezani odprti podatki (LOD) v praksi</i>
11.20 - 11.40	Iztok Kosem, Tadeja Rozman, Špela Arhar Holdt, Polonca Kocjančič, Cyprian Laskowski: <i>Šolar 2.0: nadgradnja korpusa šolskih pisnih izdelkov</i>
11.40 - 12.00	Simon Krek, Polona Gantar, Iztok Kosem, Vojko Gorjanc, Cyprian Laskowski: <i>Baza kolokacijskega slovarja slovenskega jezika</i>
12.00 - 12.20	Špela Arhar Holdt, Iztok Kosem: <i>Ohranjanje jezikovne zahtevnosti besedil pri prevajanju testov PISA</i>
11.00-12.30	Paralelna sekcija 5B / Parallel session 5B (predavalnica / room: 18)
11.00 - 11.20	Teja Goli, Eneja Osrajnik, Darja Fišer: <i>Analiza krajšanja slovenskih sporočil na družbenem omrežju Twitter</i>
11.20 - 11.40	Damjan Popič, Darja Fišer, Katja Zupan, Polona Logar: <i>Raba vejice v uporabniških spletnih vsebinah</i>
11.40 - 12.00	Matija Ogrin, Andrejka Žejn: <i>Strojno podprta kolacija slovenskih rokopisnih besedil. Variantna mesta v luči računalniških algoritmov in vizualizacij</i>
12.00 - 12.15	Matija Ogrin, Tomaž Erjavec, Jan Jona Javoršek: <i>Od znanstvenokritične izdaje do repozitorija rokopisov. Dosežki, možnosti in načrti</i>

12.30-13.30	Odmor za kosilo / Lunch break
13.30-14.40	Študentska sekcija 1 / Student session 1 (predavalnica / room: 34)
<i>13.30 - 13.50</i>	Anja Tavčar, Ines Čeligoj Pregelj, Miha Pompe: <i>Korpus študentov prevajanja MetaTrans</i>
<i>13.50 - 14.10</i>	Jure Škerl: <i>RBP pri strojnih prevajalnikih Amebis Presis, Google Translate in MT@EC</i>
<i>14.10 - 14.25</i>	Lucija Dačić: <i>Literature Reloaded: Using Databases to Explore Literary Trends</i>
<i>14.25 - 14.40</i>	Jernej Rihter: <i>Digitalna arheologija? Primer uporabe digitalnih orodij za analizo arheološkega najdišča</i>
14.40-15.00	Odmor za kavo / Coffee break
15.00-15.40	Študentska sekcija 2 / Student session 2 (predavalnica / room: 34)
<i>15.00 - 15.20</i>	Miha Helbl, Žiga Domevšček: <i>Gradnja in analiza petjezičnega korpusa podnaslovov govorov TED</i>
<i>15.20 - 15.40</i>	Katerina Pertot, Maja Petrovčič, Nika Stojan: <i>#Analiza novih komunikacijskih elementov na družbenem omrežju @Twitter</i>
15.40-16.00	Zaključek konference s podelitvijo študentske nagrade / Conference closing with best student paper award (predavalnica / room: 34)
16.00-18.00	SDJT občni zbor (predavalnica / room: 34)

Kazalo / Table of Contents

Predgovor	i
Preface	ii
Organizacijski odbor / Organising committee	iii
Programski odbor / Programme committee	iii
Vodje tematskih področij / Area chairs	iv
Člani programskega odbora / Programme committee members	v
Organizatorji / Organizers	vii
Sponsorji / Sponsors	viii
Urniki / Timetable	ix
Kazalo / Table of Contents	xiv
VABLJENI PRISPEVKI / INVITED TALKS	1
Contrasting post-editing and human translation along the dimension of term and cognate variation <i>Oliver Čulo</i>	1
Speech Is Golden – On ASR at the Service of the Danish Public Sector <i>Peter Juel Henriksen</i>	4
The Text Encoding Initiative: 30 Years of Accumulated Wisdom and Its Potential for a Bright Future <i>Laurent Romary</i>	5
The Landscape of Digital Annotation and Its Meaning <i>Niels-Oliver Walkowski</i>	6
Universal Dependencies for Slavic Languages <i>Daniel Zeman</i>	12
PRISPEVKI / PAPERS	13
Asistent – A Machine Translation System for Slovene, Serbian and Croatian <i>Mihael Arčan, Maja Popović, Paul Buitelaar</i>	13
Ohranjanje jezikovne zahtevnosti besedil pri prevajanju testov PISA <i>Špela Arhar Holdt, Iztok Kosem</i>	21
Predstavitveni portal spletnih jezikovnih virov za slovenščino <i>Špela Arhar Holdt, Kaja Dobrovoljc, Iztok Kosem</i>	27
Integrating Natural Language and Formal Analysis for Legal Documents <i>Shaun Azzopardi, Albert Gatt, Gordon J. Pace</i>	32
Organiziranje projekta vizualizacije podatkov <i>Narvika Bovcon, Jure Demšar, Aleš Vaupotič</i>	36
Razvoj učne množice za izboljšano označevanje spletnih besedil <i>Jaka Čibej, Špela Arhar Holdt, Tomaž Erjavec, Darja Fišer</i>	40

Preizkus Googlevega govornega programskega vmesnika pri samodejnem razpoznavanju govornjene slovenščine <i>Simon Dobrišek, David Čefarin, Vitomir Štruc, France Mihelič</i>	47
Povezljivost pravopisnih pravil in slovarja: sanje pravopiscev 20. stoletja <i>Helena Dobrovoljc</i>	52
Slovenska akademska besedila: prototipni korpus in načrt analiz <i>Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Nataša Logar, Milan Ojsteršek</i>	58
Sentiment Annotation of Slovene User-Generated Content <i>Darja Fišer, Jasmina Smailović, Tomaž Erjavec, Igor Mozetič, Miha Grčar</i>	65
Slovar tviterščine <i>Polona Gantar, Iza Škrjanec, Darja Fišer, Tomaž Erjavec</i>	71
Analiza krajšanja slovenskih sporočil na družbenem omrežju Twitter <i>Teja Goli, Eneja Osrajnik, Darja Fišer</i>	77
Analiza mnenj s pomočjo strojnega učenja in slovenskega leksikona sentimenta <i>Klemen Kadunc, Marko Robnik-Šikonja</i>	83
Building a Gold Standard for Temporal Entity Extraction from Medieval German Texts <i>Natalia Korchagina</i>	90
Šolar 2.0: nadgradnja korpusa šolskih pisnih izdelkov <i>Iztok Kosem, Tadeja Rozman, Špela Arhar Holdt, Polonca Kocjančič, Cyprian Laskowski</i>	95
Baza kolokacijskega slovarja slovenskega jezika <i>Simon Krek, Polona Gantar, Izток Kosem, Vojko Gorjanc, Cyprian Laskowski</i>	101
Označevanje udeleženskih vlog v učnem korpusu za slovenščino <i>Simon Krek, Polona Gantar, Kaja Dobrovoljc, Iza Škrjanec</i>	106
Priprava usporedivih korpusa za usporedbu <i>Ivana Lalli Pačelat</i>	111
Easily Accessible Language Technologies for Slovene, Croatian and Serbian <i>Nikola Ljubešić, Tomaž Erjavec, Darja Fišer, Tanja Samardžić, Maja Miličević, Filip Klubička, Filip Petkovski</i>	120
Strojno podprta kolacija slovenskih rokopisnih besedil: variantna mesta v luči računalniških algoritmov in vizualizacij <i>Matija Ogrin, Andrejka Žejn</i>	125
Popisi prebivalstva Slovenije 1830–1931: Orodje za transkribiranje historičnih demografskih podatkov <i>Andrej Pančur</i>	133
Označevanje zbirke zapisnikov sej slovenskega parlamenta s smernicami TEI <i>Andrej Pančur</i>	142
Raba vejice v uporabniških spletnih vsebinah <i>Damjan Popič, Darja Fišer, Katja Zupan, Polona Logar</i>	149

EAGLE – Medomrežje Europeana antične grške in latinske epigrafike.	
Digitalni dostop do antičnih napisnih spomenikov	
<i>Anja Ragolič</i>	154
ARIADNE: povezani odprti podatki (LOD) v praksi	
<i>Benjamin Štular, Franco Niccolucci, Julian Richards</i>	158
Digital Video in Digital Humanities Methodology: A Case Study	
<i>Aleš Vaupotič, Marco Buziol, Narvika Bovcon</i>	164
Digitalna humanistika v šoli	
<i>Maja Vičič Krabonja</i>	170
Generiranje kritičnih prepisov s strojnim prevajanjem na ravni znakov	
<i>Katja Zupan, Tomaž Erjavec</i>	175
Sintetizator govora za slovenščino eBralec	
<i>Jerneja Žganec Gros, Boštjan Vesnicer, Simon Rozman, Peter Holozan, Tomaž Šef</i>	180
Razpoznavanje tekočega govora v slovenščini z bazo predavanj SI TEDx-UM	
<i>Andrej Žgank, Darinka Verdonik, Mirjam Sepesy Maučec</i>	186
POVZETKI / ABSTRACTS	190
Pretvorba korpusa ssj500k v Univerzalno odvisnostno drevesnico za slovenščino	
<i>Kaja Dobrovoljc, Tomaž Erjavec, Simon Krek</i>	190
What is critical in digital humanities?	
<i>Mario Hibert</i>	193
Asian Language Teaching and Learning - The Influence of Technology on Students' Skills in SL Classroom	
<i>Marijana Janjić, Sara Librenjak, Kristina Kocijan</i>	196
Zbiva in EWD, spletni orodji za arheološke raziskave	
<i>Bojan Kastelic, Mateja Belak, Andrej Pleterski, Benjamin Štular, Miran Erič</i>	198
Nadgradnja korpusov Gigafida, Kres, ccGigafida in ccKres	
<i>Simon Krek, Polona Gantar, Špela Arhar Holdt, Vojko Gorjanc</i>	200
Digitalizacija in uporaba digitalnega etnomuzikološkega zvočnega gradiva – izkušnje Glasbenonarodopisnega inštituta ZRC SAZU	
<i>Drago Kunej</i>	204
Trirazsežno dokumentiranje v službi varovanja nepremične kulturne dediščine	
<i>Aleš Lazar, Sonja Ifko</i>	205
Analysing Spatial Distribution of Linguistic Variables in Geocoded Tweets from Croatia, Bosnia, Montenegro and Serbia	
<i>Nikola Ljubešić, Tanja Samardžić, Maja Miličević</i>	208
The First World War on the Web - The Case of Serbia	
<i>Slobodan Mandić</i>	210
Od znanstvenokritične izdaje do repozitorija rokopisov: Dosežki, možnosti in načrti	
<i>Matija Ogrin, Tomaž Erjavec, Jan Jona Javoršek</i>	214

The Use of Semantic Word Classes in Document Classification <i>Stevan Ostrogonac, Branislav Popović, Milan Sečujski</i>	216
Dolgotrajno ohranjanje raziskovalnih podatkov v manjših raziskovalnih infrastrukturah: Uporaba odprtokodne aplikacije Archivemata <i>Andrej Pančur, Bogomir Rožman</i>	218
Razvoj aplikacije za spodbujanje trajnostne mobilnosti <i>Dan Podjed, Saša Babič, Tatiana Bajuk Senčar, Alenka Bezjak Mlakar, Gregor Burger, Jurij Fikfak, Jože Guna, Marko Maver, Matevž Pogačnik, Emilija Stojmenova Duh, Uroš Žolnir</i>	220
Luščenje in jezikoslovna analiza kolokacij iz korpusa Šolar <i>Tadeja Rozman, Špela Arhar Holdt, Senja Pollak, Iztok Kosem</i>	222
Historična topografija Slovenije <i>Miha Seručnik</i>	225
Language Technologies in Humanities: Computational Semantic Analysis in Folkloristics <i>Gregor Strle, Matija Marolt</i>	227
Slikovna retrospektiva porušenega Breginja in analiza pokrajinskih sprememb <i>Tatjana Veljanovski, Žiga Kokalj</i>	230
ŠTUDENTSKI PRISPEVKI / STUDENT PAPERS	233
Gradnja in analiza petjezičnega korpusa podnaslovov govorov TED <i>Miha Helbl, Žiga Domevšček</i>	233
#Analiza novih komunikacijskih elementov na družbenem omrežju @Twitter <i>Katerina Pertot, Maja Petrovčič, Nika Strojjan</i>	239
Razdvoumljanje besednega pomena pri strojnih prevajalnikih Amebis Presis, Google Translate in MT@EC <i>Jure Škerl</i>	245
Korpus študentov prevajanja MetaTrans <i>Anja Tavčar, Ines Čeligoj Pregelj, Miha Pompe</i>	250
ŠTUDENTSKI POVZETKI / STUDENT ABSTRACTS	256
Literature Reloaded: using databases to explore literary trends <i>Lucija Dačić</i>	256
Digitalna arheologija? Primer uporabe digitalnih orodij za analizo arheološkega najdišča <i>Jernej Rihter</i>	258