

KONFERENCA JEZIKOVNE TEHNOLOGIJE IN DIGITALNA HUMANISTIKA 2020

Katja MEDEN

Institut Jožef Stefan

Meden, K. (2020): Konferenca Jezikovne tehnologije in digitalna humanistika 2020. Slovenščina 2.0, 8(1): 113–119.

DOI: <https://doi.org/10.4312/slo2.0.2020.1.113-119>

Konferenca Jezikovne tehnologije in digitalna humanistika 2020, ki jo skupaj z Inštitutom za novejšo zgodovino,¹ Centrom za jezikovne vire in tehnologije Univerze v Ljubljani (CJVT UL)² ter raziskovalnima infrastrukturama CLARIN-SI³ in DARIAH-SI⁴ organizira Slovensko društvo za jezikovne tehnologije (SDJT),⁵ je letos potekala 24. in 25. septembra 2020, že tretjo multidisciplinarno izvedbo konference pa je podprl CLARIN ERIC.⁶ Konferenca, ki se lahko pohvali z več kot 20-letno tradicijo delovanja, je leta 2016 v svoj program vključila tudi področje digitalne humanistike in s tem postala pomemben povezovalni člen med omenjenima disciplinama.

Letošnja konferenca bi morala potekati na Inštitutu za novejšo zgodovino v Ljubljani, vendar je bila zaradi pandemije Covid-19 preseljena v virtualno okolje. Digitalna izvedba je predstavljala odmik od tradicionalne organizacije konference, hkrati pa je prinesla tudi nove izzive. Konferenca je tradicionalno potekala dvodnevno, vendar pa se v nasprotju s predlansko ni delila po jezikih, v katerih so bili prispevki predstavljeni, temveč izključno po tematskih sklopih. Prispevki v posameznih sekcijah so bili kot vedno do sedaj razdeljeni na polne in razširjene. Posamezni študentski prispevki so bili v letošnji izvedbi

1 <https://www.inz.si/>

2 <https://www.cjvt.si/>

3 <https://www.clarin.si/info/about/>

4 <http://www.dariah.si/>

5 <http://www.sdjt.si/wp/>

6 <https://www.clarin.eu/>

umeščeni znotraj tematsko ustreznih sekcij in ne v ločenem študentskem panelu. Ker se je v letošnjem letu zaradi globalne epidemije konkretno spremenil način dela in znanstvenega komuniciranja, so avtorji posameznih prispevkov svoje predstavitve predhodno posneli, te pa so bile na spletni strani na voljo že pred začetkom konference. To je omogočilo dinamično izvedbo konferenčnih debat, saj so si udeleženci posnetke in zbornik prispevkov ogledali že pred samo konferenco, v okviru panela pa so nato avtorji predstavili le kratek povzetek svojih predavanj, ki jim je sledila razprava.

Konferenco je v sklopu prvega vabljenega predavanja, ki ga je moderiral Filip Dobranić, odprla Sara Tonelli, vodja raziskovalne skupine Digitalna humanistika pri Fondaciji Bruna Kesslerja v Trentu in izredna profesorica na Oddelku za psihologijo in kognitivne znanosti na Univerzi v Trentu. V svojem predavanju z naslovom *Abusive Language Detection: Too much Digital, not enough Humanities?* je predavateljica ustvarila pregled sodobnih pristopov za prepoznavanje nasilnega govora, poudarila vlogo in razumevanje jezika, ki ga uporabljajo različne spletne skupnosti, ter predstavila trenutne raziskave za boljše razumevanje tega fenomena.

Vabljenemu predavanju je sledila prva sekcija na temo jezikovnih in govornih tehnologij, ki jo je povezovala Tanja Samardžić. Med polnimi prispevki lahko izpostavimo predstavitev prispevka Anke Supej, Mateja Ulčarja, Marka Robnika Šikonje in Senje Pollak, v katerem so primerjali spolno pristranost modelov (oziroma njihovih vložitev) z različnimi konfiguracijami in pristopi k računanju analogij. Med predstavitvami razširjenih povzетkov je Darinka Verdonik v imenu soavtorjev predstavila razvoj in delovanje raziskovalne infrastrukture projekta RI-SI CLARIN, v študentskem prispevku pa je Andraž Pelicon predstavil zaznavanje sentimenta v novicah s pomočjo globokih nevronske mreže.

Naslednjo sekcijo, ki je obravnavala jezikovne vire, je moderirala Darinka Verdonik. Vsebovala je predstavitve štirih polnih prispevkov, treh razširjenih povzетkov in študentski prispevek. Simon Krek (s soavtorji) je predstavil najnovejše razvojne dosežke in vsebino učnega korpusa ssj500k,⁷ največje in najpogosteje uporabljene odprtokodne zbirke podatkov za obdelavo slovenskega

7 <https://www.clarin.si/repository/xmlui/handle/11356/1210>

jezika. Dolores Lemmenmeier-Batinić je izpostavila problem pomanjkanja javno dostopnih virov za srbski jezik, opisala do sedaj ustvarjene korpuse in metode za pretvorbo korpusnih podatkov v standardizirani XML format. Nazadnje je sekcijo s predstavitevijo študentskega prispevka o metodoloških izhodiščih, razvoju in vodilih pri označevanju metaforičnih besed korpusa metafor *KOMET 1.0* sklenila Špela Antloga. Prvi dan konference je zaključil poseben panel projekta *RSDO – Razvoj slovenščine v digitalnem okolju*, v okviru katerega so vodje posameznih delovnih sklopov predstavili glavne in vmesne cilje ter naloge, ki jih bodo uresničevali v okviru projekta.

Uvod v drugi dan konference, ki je bil obarvan s prepletanjem jezikovnih tehnologij, predvsem korpusno analizo in digitalno humanistiko, je ponudil digitalni zgodovinar Kaspar Beelen z Inštituta Alan Turing, ki raziskuje uporabo strojnega učenja v humanističnih raziskavah. V vabljenem predavanju z naslovom *Speaking on behalf of others: Why the digital humanities should care about parliamentary data* je izpostavil pomen in vlogo parlamentarnih podatkov, ki omogočajo vpogled v jezik ter svetovni nazor poslancev in volivcev, ki naj bi jih ti zastopali, s tem pa zagotavljajo podrobno pričevanje o skoraj vsaki temi, ki je bila kdaj koli predmet javne razprave.

V tretji sekciji, ki jo je moderirala Kristina Štrkalj Despot, so bili predstavljeni prispevki s področja korpusne analize. Kristina Pahor de Maiti je predstavila študijo morfosintaktičnih značilnosti komentarjev na Facebooku, da bi prepoznali tiste, ki so vidni v družbeno nesprejemljivem diskurzu (SUD). Jakob Lenardič in Darja Fišer sta analizirala epistemske modalne prislove v 100-milijonskem korpusu slovenskih doktorskih disertacij (korpus KAS). Študentska prispevka sta predstavljala Zoran Fijavž in Eva Trivunović. Zoran Fijavž je v svojem prispevku raziskoval vpliv videovsebin na prisotnost družbeno neperimerne diskurza, kar je detektiral iz nabora komentarjev na Facebooku, povezanih z LGBT, ki so izvirali iz glavnih virov novic v hrvaščini. Eva Trivunović pa je predstavila (iz)biblijske frazeme, njihove variante, prenovitvene in neprenovitvene modifikacije v korpusih *Gigafida 2.0*, *Janes* in *slWaC*.

Zadnjo sekcijo, v kateri sta se združili digitalna humanistika in pedagogika, je moderiral Miran Hladnik, vsebovala pa je predstavitve treh polnih prispevkov, dveh razširjenih povzetkov ter eno predstavitev študentskega prispevka. Katja Meden in Ana Cvek sta predstavili tehnično in vsebinsko nadgradnjo

Zgodovinskega indeksa citiranosti, citatnega indeksa za sistematično popisovanje citiranosti del na področju zgodovinopisja. Andrej Pančur je predstavil najnovejšo pridobitev med demografsko viktimološkimi raziskavami, digitalno bazo podatkov (vojaških) žrtev prve svetovne vojne z ozemlja današnje Republike Slovenije, ki je nastala na podlagi sodelovanja med različnimi raziskovalnimi in kulturnimi ustanovami ter nekaterimi posamezniki. Nazadnje je Magdalena Schlintl v študentskem prispevku predstavila delo z digitalnimi orodji pri izobraževanju učiteljev na primeru poučevanje-učenje-laboratorij.

Konferenca se je zaključila s podelitvijo nagrade za najboljši študentski prispevek, ki jo je prejel Zoran Fijavž za prispevek *Ambivalence of Queer Visibility in Video-Based Social Media Content*. Ker je bilo med posnetimi predstavitev kar nekaj izstopajočih, se je programski odbor konference odločil v zaključnem delu podeliti tudi nagrado za najboljšo posneto predstavitev po izboru publike. Nagrado je prejela Špela Antloga za predstavitev prispevka *Korpus metafor KOMET 1.0*.

Ne glede na tehnične in organizacijske izzive, ki sta jih prinesla pandemija in virtualno okolje, je bila Konferenca jezikovne tehnologije in digitalna humanistika 2020 uspešno izpeljana, tako z vsebinskega kot tudi z organizacijskega vidika. Ravno zaradi digitalne izvedbe so na spletni strani konference⁸ na voljo ne le zbornik prispevkov, temveč tudi vse predstavitve in posnetki diskusij. Prispevki, predstavljeni na konferenci, so omogočili vpogled v nove metode, aplikacije, nadgradnjo in razvoj raziskovalnih polj. Razprave, ki so potekale med predstavitev, pa so udeležence spodbudile k nadaljevanju raziskav in jim ponudile vpogled v interdisciplinarnost različnih tematskih podpolj jezikovnih tehnologij in digitalne humanistike. Preprosto povedano, konferenca je prinesla priložnost za tesnejše sodelovanje s sorodnimi področji ter postavila nove gradnike pri premoščanju vrzeli med jezikovnimi tehnologijami in digitalno humanistiko. Izbrani prispevki bodo v obliki izvirnih znanstvenih člankov na voljo v naslednji tematski številki revije *Slovenščina 2.0*.

8 Konferenca Jezikovne tehnologije in digitalna humanistika: <http://www.sdjt.si/wp/dogodki/konference/jtdh-2020/>.

LITERATURA

- Antloga, Š. (2020). Korpus metafor KOMET 1.0. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 167–170). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Student_Antloga_Korpus-metafor-KOMET-1.0.pdf
- Beelen, K. (2020). Speaking on behalf of others: Why the digital humanities should care about parliamentary data. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika: 2*. Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Invited-Lecture_Beelen_Speaking-on-behalf-of-others-Why-the-digital-humanities-should-care-about-parliamentary-data.pdf
- Fijavž, Z. (2020). Ambivalence of LGBT Visibility in Croatian Video-Based Social Media Content. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 144–148). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Student_Fijavz_Ambivalence-of-LGBT-Visibility-in-Croatian-Video-Based-Social-Media-Content.pdf
- Konferenca Jezikovne tehnologije in digitalna humanistika. Dostopno prek <http://www.sdjt.si/wp/dogodki/konferenca/jtdh-2020/>
- Krek, S., Erjavec, T., Dobrovoljc, K., Gantar, P., Arhar Holdt, Š., Čibej, J. in Brank, J. (2020). The ssj500k Training Corpus for Slovene Language Processing. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 24–33). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Krek-et-al_The-ssj500k-Training-Corpus-for-Slovene-Language-Processing.pdf
- Lemmenmeier-Batinić, D., Ljubešić N. in Samardžić, T. (2020). XML-Encoding of a spoken Serbian corpus targeting forms of address. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 127–130). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Lemmenmeier-Batinic-et-al_XML-Encoding-of-a-spoken-Serbian-corpus-targeting-forms-of-address.pdf
- Lenardič, J. in Fišer, D. (2020). Epistemic modal adverbs in Slovenian academic discourse. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 34–41).

- Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Lenardic-et-al_Epistemic-modal-adverbs-in-Slovenian-academic-discourse.pdf
- Meden, K. in Cvek, A. (2020). Nadgradnja Zgodovinarskega indeksa citiranosti. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 42–47). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Meden-et-al_Nadgradnja-Zgodovinarskega-indeksa-citiranosti.pdf
- Pahor de Maiti, K., Fišer, D., Ljubešić, N. in Erjavec, T. (2020). Grammatical Footprint of Socially Unacceptable Facebook Comments. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 48–57). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_PahordeMaiti-et-al_Grammatical-Footprint-of-Socially-Unacceptable-Facebook-Comments.pdf
- Pančur, A., Blaj Hribar, N., Ojsteršek, M. in Šorn, M. (2020). Projekt Vojaške žrtve prve svetovne vojne na Slovenskem. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 136–140). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Pancur-et-al_Projekt-Vojaske-zrtve-prve-svetovne-vojne-na-Slovenskem.pdf
- Schlintl, M., Pawluch, K., Rader, M. in Novak-Geiger, V. (2020). Working with Digital Devices in Teacher Training Using the Example of the Teaching-Learning-Lab. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 171–174). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Student_Schlintl-et-al_Working-with-Digital-Devices-in-Teacher-Training-Using-the-Example-of-the-Teaching-Learning-Lab.pdf
- Supej, A., Ulčar, M., Robnik-Šikonja, M. in Pollak, S. (2020): Primerjava slovenskih besednih vektorskih vložitev z vidika spola na analogijah poklicev. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 93–100). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Supej-et-al_Primerjava-slovenskih-besednih-vektorskih-vlozitev-z-vidika-spola-na-analogijah-poklicev.pdf

- Tonelli, S. (2020). Abusive Language Detection: Too much Digital, not enough Humanities?. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika: 1*. Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Invited-Lecture_Tonelli_Abusive-Language-Detection-Too-much-Digital-not-enough-Humanities.pdf
- Trivunović, E. (2020). Variante in modifikacije (iz)biblijskih frazemov. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 158–166). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek [http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Student-Trivunovic_Variante-in-modifikacije-\(iz\)biblijskih-frazemov.pdf](http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Student-Trivunovic_Variante-in-modifikacije-(iz)biblijskih-frazemov.pdf)
- Verdonik, D., Rojc, M., Mlakar, I., Zimšek, D., Kačič, Z., Ojsteršek, M. in Erjavec, T. (2020). Raziskovalna infrastruktura projekta RI-SI CLARIN. V D. Fišer in T. Erjavec (ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 141–143). Ljubljana: Inštitut za novejšo zgodovino. Dostopno prek http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Verdonik-et-al_Raziskovalna-infrastruktura-projekta-RI-SI-CLARIN.pdf