

Datum: 26. 11. 2025

Vabilo na javno predstavitev rezultatov Centra za jezikovne vire in tehnologije Univerze v Ljubljani

10. december 2025, ob 10:30

Fakulteta za računalništvo in informatiko (Diplomska soba, 1. nadstropje), Večna pot 113, 1000 Ljubljana

Vljudno vabljeni na javno predstavitev rezultatov Centra za jezikovne vire in tehnologije Univerze v Ljubljani, ki bo potekala v sredo, 10. decembra, ob 10:30 v Diplomski sobi Fakultete za računalništvo in informatiko.

Center za jezikovne vire in tehnologije je interdisciplinarna enota Univerze v Ljubljani. Naše poslanstvo je omogočiti uspešen razvoj slovenščine v digitalni dobi. Pri tem sledimo načelom, da morajo biti vsi izdelki vsebinsko in metodološko sodobni, uporabniku prijazni in odprto dostopni.

Na dogodku bomo predstavili:

- Novi različici Referenčnega korpusa pisne slovenščine Gigafida in korpusa Trendi.
- Rezultate projekta PODVIG in drugih projektov, v okviru katerih so nastale nove različice digitalnih slovarjev (Sopomenke, Kolokacije, SLOGOST, Veliki slovensko-madžarski slovar).
- Nov projekt občanske znanosti s Slovarjem sopomenk sodobne slovenščine.
- Dve novi jezikovni igri na portalu CJVT Igre: Ugibanko in Kombinator.
- Rezultate projekta SPOT o značilnostih slovenskega govora.
- Orodje za strojno prevajanje slovenskih predavanj v angleščino Online Notes.
- GaMS 3 in Slovensko pogovorno areno.
- Raziskavo motivov v pravljicah o Pepelki s pomočjo umetne inteligence.
- Napoved izobraževanj v okviru Centra odličnosti za umetno inteligenco v digitalni humanistiki.
- Besede leta 2025.

Več o posamezni temi lahko preberete v nadaljevanju dokumenta. Predstavitev bo trajala približno uro in pol. Po predstavitvi vabimo na druženje s pogostitvijo.

Veselimo se srečanja z vami!

Prosimo, da do petka, 5. decembra, svojo udeležbo potrdite na [povezavi](#).

Zbirka slovenskih besedil Gigafida 2.2 in spremljevalni korpus Trendi

[Korpus Gigafida](#) je referenčni korpus pisne slovenščine: obsežna, premišljeno zasnovana zbirka besedil, ki je ključna za raziskave jezika, jezikovni opis (slovarje, slovnice), pripravo učnih gradiv in razvoj raznovrstnih jezikovnih tehnologij ter strojnih postopkov za procesiranje slovenskega jezika. Gigafida vsebuje časopise, revije, določena spletna besedila, leposlovje, stvarno literaturo, šolske učbenike in podobna besedila. Gigafida je osnovni jezikovni vir za izdelke CJVT, kot so odzivni slovarji in leksikoni, podatkovne baze in portali. Na dogodku bomo predstavili novo različico korpusa.

[Spremljevalni korpus Trendi](#) vsebuje besedila z več kot 100 medijskih spletnih strani in je prvi spremljevalni korpus za slovenščino. Korpus posodabljam mesečno in vsebuje besedila od leta 2019 do danes. Njegov glavni namen je zagotavljanje vpogleda v aktualno rabo jezika – spremljanje rabe besed in besednih zvez, pojavljanja novih besed in pomenov. Na dogodku bomo predstavili novo različico spremljevalnega korpusa Trendi, ki je že presegel milijardo besed.

Projekt PODVIG

V okviru projekta Podatkovna dopolnitev in igrifikacija slovarskih virov (PODVIG) smo nagradili podatke v [Slovarju sopomenk sodobne slovenščine](#), v [Kolokacijskem slovarju sodobne slovenščine](#) in v Slovarju za govorce slovenščine kot drugega tujega jezika [SloGost](#). Na dogodku bomo predstavili nove različice slovarjev.

Projekt občanske znanosti s Slovarjem sopomenk sodobne slovenščine

Slovar sopomenk sodobne slovenščine je eno izmed najbolj priljubljenih orodij Centra za jezikovne vire in tehnologije. Pri njegovem razvoju je sodelovalo več kot 1300 občanov in občanov, ki so prispevali več kot 75.000 predlogov sopomenk in protipomenk. Zahvaljujoč financiranju Javne agencije za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS), bomo prispevke občanov in občanov prvič leksikografsko validirali in vključili v Digitalno slovarsko bazo za slovenščino.

Novi jezikovni igri na portalu CJVT Igre: Ugibanka in Kombinator

[CJVT Igre](#) so spletni portal z raznolikimi jezikovnimi igrami. Marca 2025 smo predstavili prve tri igre – Besedolov, Cvetko in Vezalko. Na portal bomo v zadnjem delu leta dodali še dve novi igri: Ugibanko in Kombinator. Portal ponuja sproščujoč in zabaven način za stik s sodobno slovenščino ter spodbuja bogatenje jezikovnih zmožnosti v slovenskem jeziku.

Rezultati projekta SPOT

Projekt SPOT (Na drevesnici temelječ pristop k raziskavam govorjene slovenščine) se osredotoča na opis skladenjskih značilnosti slovenskega govora s pomočjo nove metode na temelju skladenjsko razčlenjenih zbirk pisnih in govorjenih besedil (t. i. drevesnic). Na dogodku bomo predstavili rezultate raziskav o govorjeni slovenščini.

Orodje za strojno prevajanje slovenskih predavanj v angleščino Online Notes

V okviru projekta [Razvoj sistema za avtomatizirano prevajanje slovenskih predavanj v tuje jezike \(Online Notes\)](#) smo razvili orodje, ki omogoča sprotno strojno razpoznavo govora in prevajanje slovenskih predavanj v tuje jezike za študente, ki prihajajo iz tujine. Poleg tega omogoča spremljanje učnega procesa tudi tistim, ki zaradi senzoričnih in fizičnih ovir težje spremljajo predavanja. Sistem je bil razvit z mislijo na dilemo glede jezika izvajanja predavanj

na Univerzi v Ljubljani. Strateški cilj Univerze v Ljubljani je postati mednarodno prepoznavna in za tuje študente ter profesorje odprta ustanova. Ohranjanje slovenskega jezika na vseh ravneh izobraževanja je cilj vseh, saj je jezik pomemben del identitete naroda. Sistem Online Notes pomaga uresničiti oba cilja.

GaMS 3 in Slovenska pogovorna arena

V okviru projekta PoVeJMo gradimo veliki generativni jezikovni model za slovenščino, ki smo ga poimenovali GaMS. Na dogodku bomo predstavili novo različico velikega jezikovnega modela za slovenščino GaMS 3. Veliki jezikovni model smo razvili s podporo NVIDIA Academic Grant Program.

Za razvoj jezikovnega modela potrebujemo besedila v obsegu 40 milijard besed. Na dogodku bomo predstavili število zbranih besed v okviru nacionalne zbiralne akcije. Predstavili bomo tudi [Slovensko pogovorno areno](#), na kateri lahko ocenite odgovore različnih velikih jezikovnih modelov. Cilj je dobiti uporabniško evalvacijo velikih jezikovnih modelov za slovenščino in učno množico za prilagajanje slovenskih modelov.

Pepelka in umetna inteligenca

Digitalna humanistika združuje humanistične vede z računalniškimi orodji in metodami. Umetna inteligenca lahko bistveno izboljša raziskave na tem področju, saj omogoča hitrejšo analizo, transkripcijo in prevajanje podatkov ter njihovo shranjevanje v digitalnih arhivih. Algoritmi umetne inteligence so zmožni v velikih podatkovnih zbirkah prepoznati povezave, vzorce in trende, ki raziskovalcem prek uporabe tradicionalnih metod ostajajo skriti. S pomočjo umetne inteligence smo raziskali motive v veliki zbirki različnih pravljič o Pepelki. Na dogodku bomo predstavili rezultate raziskave.

Napoved izobraževanj Centra odličnosti za umetno inteligenco v digitalni humanistiki

Evropski Center odličnosti za umetno inteligenco v digitalni humanistiki raziskovalcem s področja humanistike zagotavlja podporo pri uporabi orodij umetne inteligence, strojnega učenja in analize podatkov. Na voljo je strokovna pomoč pri razvoju novih pristopov, prilagojenih humanističnim področjem, kot so jezikoslovje, zgodovina, medijske študije, folkloristika, novinarstvo in filozofija. Na dogodku bomo predstavili izobraževanja, ki jih bomo v letu 2026 izvedli za raziskovalce in raziskovalke s področja humanistike in družboslovja.

Razglasitev besed leta 2025

Na Centru za jezikovne vire in tehnologije že od konca leta 2020 vsak december objavimo besede, ki so zaznamovale iztekajoče se leto, predvsem jezik spletnih medijev. Na dogodku bomo razglasili besede leta 2025. Besede leta niso izbrane naključno. Na mesečni ravni zbiramo besedila iz 106 slovenskih množičnih medijev. Nato s pomočjo statističnih metod izračunamo količnik pojavljanja določene besede glede na podatke iz preteklih obdobj. Posledično nam izbor besed leta veliko pove tudi o slovenski medijski krajini in o temah, o katerih so največ pisali slovenski mediji.