

HR-CLARIN: repozitorij europske istraživačke infrastrukture CLARIN ERIC u Hrvatskoj

Marko Tadić, Daša Farkaš

Sveučilište u Zagrebu, Filozofski fakultet
{mtadic, dfarkas}@ffzg.unizg.hr

Četrnaesti festival hrvatskih digitalizacijskih projekata

NSK, Zagreb, 2025-05-09



Pregled izlaganja

- europska istraživačka infrastruktura CLARIN ERIC
- hrvatski konzorcij HR-CLARIN
- repozitorij HR-CLARIN-a
- kampanja prikupljanja metapodataka o jezičnim resursima

Što je CLARIN ERIC?

- CLARIN
 - pokrata za *Common LAnguage Resources and technology INfrastructure (for Humanities and Social Sciences)*
 - zajednički jezični resursi i tehnološka infrastruktura za humanističke i društvene znanosti (HiDZ)
- europska istraživačka infrastruktura unutar ESFRI-ja
 - za HiDZ: DARIAH, CLARIN, ESS, CESSDA, SHARE, (OPERAS, EHRI u nastajanju)
 - www.clarin.eu
- nastala po zaključenju pripremnoga FP7 projekta CLARIN
 - trajao 2008. - 2011.

Što je CLARIN ERIC?

- osnovan 2012-04-18, 9 (zemalja/udruga)-osnivačica
 - Austrija, Bugarska, Češka, Danska, Estonija, Nizozemska, Nizozemska jezična unija, Njemačka, Poljska
 - Hrvatska nije bila među njima premda je potpisala Memorandum o razumijevanju 2011-12
 - MZO i MVEP nisu u 2012. dostavili potvrdu da RH priznaje ERIC kao nadnacionalnu pravnu osobu
 - 5-godišnja članarina za RH: oko 70.000 EUR
- svaka punopravna članica preuzela je obvezu financiranja nacionalnoga CLARIN konzorcija
 - izgrađuje infrastrukturu na nacionalnoj razini
 - povezuje tu infrastrukturu s CLARIN ERIC-om na EU razini

Misija projekta CLARIN

- što?
 - stvoriti istraživačku infrastrukturu koja omogućuje uporabu jezičnih resursa (JR) i jezičnih tehnologija (JT) istraživačima ih svih znanosti, a osobito HiDZ
- kako?
 - ujediniti postojeće digitalne repozitorije jezičnih podataka u **europsku federaciju digitalnih repozitorija**
 - ciljana skupina korisnika: **znanstvenici iz HiDZ**
 - pristup: **jedinstven i jednoobrazan pristup** svim repozitorijima
 - sredstva: **jezičnotehnološki alati** kao *web*-servisi za pretragu, rukovanje, obogaćivanje i korištenje podataka iz tih repozitorija
 - pokrivenost: **svi** su **jezici** jednako bitni
 - razmještaj: mogu sudjelovati **sve države EU i države kandidati**

HR-CLARIN

- Hrvatska postaje članica CLARIN ERIC-a tek 2018.
- konzorcij HR-CLARIN
 - Sveučilište u Zagrebu, Filozofski fakultet (FFZG), koordinator
 - Sveučilište u Zagrebu, Sveučilišni računski centar (SRCE)
 - Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva (FER)
 - Institut za hrvatski jezik (IHJ)
 - Staroslavenski institut (STIN)
 - Nacionalna i sveučilišna knjižnica (NSK)
 - Hrvatska akademija znanosti i umjetnosti – Knjižnica (HAZU)
 - Hrvatsko društvo za jezične tehnologije (HDJT)
- konzorcij je otvoren za druge ustanove koje proizvode JR-ove

HR-CLARIN 2

- otvorio pristup hrvatskim znanstvenicima u CLARIN ERIC
- prioriteti konzorcija HR-CLARIN
 - učiniti hrvatske jezične resurse, alate i servise dostupnima unutar zajednice CLARIN ERIC, ali omogućiti pristup i hrvatskim istraživačima (aai@edu.hr)
 - ustanoviti i kontinuirano nadopunjavati repozitorij HR-CLARIN-a
 - ustanoviti bar jedan HR-CLARIN K-centar u srednjoročnom razdoblju
 - aktivno sudjelovati u projektima i aktivnostima CLARIN ERIC-a
 - podizati svijest o CLARIN ERIC-u među istraživačima iz (digitalnih) HiDZ na nacionalnoj razini i šire
- <https://www.clarin.hr>

Repozitorij HR-CLARIN-a

- uspostavljen nakon što su zadovoljeni svi tehnički preduvjeti:
 - punopravno članstvo Hrvatske u CLARIN ERIC-u
 - uspostavljen hrvatski konzorcij HR-CLARIN
 - prikupljeni metapodatci o odabranim jezičnim resursima
- hrvatskim znanstvenicima otvoren pristup pohrani i dijeljenju jezičnih resursa
- svaki jezični resurs dobiva jedinstveni trajni identifikator (PID) i može se citirati kao stalni izvor jezičnih podataka
- <https://repository.clarin.hr>

Repozitorij HR-CLARIN-a unutar federacije

- svi repozitoriji unutar CLARIN-ove infrastrukture komuniciraju sa središnjim agregatorom
 - obuhvaća metapodatke svih resursa iz cijele federacije repozitorija
- objedinjena pretraga s pomoću zajednički definiranih i standardiziranih metapodatkovnih shema za različite vrste resursa
- semantička interoperabilnost
- Virtualni jezični opservatorij (*Virtual Language Observatory*, VLO) - objedinjeno sučelje za pretragu jezičnih resursa
 - <https://vlo.clarin.eu>

Kampanja prikupljanja (meta)podataka

- prikupljamo (meta)podatke o jezičnim resursima koje ste spremni podijeliti putem HR-CLARIN-ova repozitorija
 - učiniti ih vidljivima u europskoj istraživačkoj infrastrukturi i globalno
 - otvoriti mogućnosti za suradnju i umrežavanje
- <https://bit.ly/HR-CLARIN-prijava>



Croatina

- CLARIN-ov centar znanja (tzv. *K-centre*) za hrvatski jezik
 - osnovan 2023., sudjeluju FFZG i IHJ
 - pruža relevantna znanja o hrvatskom jeziku i promiče korištenje jezičnih tehnologija za hrvatski jezik
 - korisnicima daje potporu (*helpdesk*) pružajući
 - relevantne informacije o temama vezanim uz hrvatski jezik
 - savjete korisnicima o izgradnji i pohrani vlastitih jezičnih resursa
 - pomoć s pristupom podacima i alatima te s uporabom jezičnih resursa
 - preporuke o širokom spektru tema vezanih uz jezične tehnologije, ponajprije za hrvatski jezik
- www.clarin.hr/hr/centri/croatina/

Zahvaljujemo na pozornosti.

Konzorcij HR-CLARIN podupire



Ministarstvo znanosti, obrazovanja i mladih Republike Hrvatske.