

Ljudi i računala otključavaju skrivena blaga prošlosti – Projekt EnrichEuropeana+

Sonja Galina

Partneri

Državni arhiv u Zagrebu sudjeluje u projektu „Enriching Europeana through citizen science and artificial intelligence - unlocking the 19th century (EnrichEuropeana+)“ zajedno s osam drugih partnera: AIT Austrian Institute of Technology GmbH iz Austrije, koji je ujedno i voditelj projekta, Stichting Europeana iz Nizozemske, Facts & Files Historisches Forschungsinstitut Berlin Drauschke - Schreiber Partnerschaftsgesellschaft iz Njemačke, Instytut Chemii Bioorganicznej Polskiej Akademii Nauk iz Poljske, READ-COOP SCE iz Austrije, The Provost, Fellows, Foundation Scholars and the other members of Board, of the College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin iz Irske, Dublin City Council iz Irske te Uniwersytet Wrocławski iz Poljske.

Cilj i zadatak projekta

Projekt je započeo 1. travnja 2021. godine, a trajat će do 30. rujna 2022. Njime se kombinira *citizen science* kampanje i umjetnu inteligenciju (AI) u olakšavanju pristupa rukom pisanim dokumentima 19. stoljeća, ne samo istraživačima nego i studentima, povjesničarima amaterima i široj javnosti.

Znatan dio arhivskog gradiva iz razdoblja 19. stoljeća koje se čuva u arhivima i drugim baštinskim ustanovama rijetko se koristi. Jedna od prepreka je rukom pisani tekst koji zahtjeva velik angažman, trud i iskustvo prilikom upotrebe. Ovaj projekt nastoji riješiti taj problem te će tijekom njegova trajanja partneri, koji su ujedno i davatelji sadržaja, digitalizirati važne, rukom pisane dokumente i zbirke koje će se potom transkribirati uz pomoć umjetne inteligencije i šire javnosti uključene u *citizen science* aktivnosti.

Projekt se oslanja na platformu Europeana koja potiče baštinske ustanove u njihovoj digitalizaciji i transformaciji pružajući im jedinstveni zajednički portal za predstavljanje kulturne baštine iz njihovih spremišta. Također nastavlja razvoj ranijih projekata koji su razvijali platformu Europeane, točnije Transcribathon platformu, koja služi kao crowdsourcing platforma za transkripcije na Europeani te Transkribus, automatski alat za transkripciju.

Projekt ima tri glavna elementa:

1. agregacija više od 100.000 novih dokumenata u Europeanu pomoću postojećih nacionalnih agregatora;
2. razvoj inovativnih AI (*artificial intelligence*) rješenja za automatizaciju transkripcije rukopisnih tekstova i semantičko obogaćivanje istih;
3. organiziranje *citizen science crowdsourcing* kampanja radi poticanja korištenja Transcribathon platforme za transkripciju, prevođenje, dodavanje semantičkih obogaćenja i povećanja dostupnosti rukom pisanih dokumenata koji su obilježili povijesne događaje i opisuju društvenu transformaciju Europe u 19. stoljeću.

Projekt je podijeljen na 5 aktivnosti:

1. *Content aggregation & data exchange Infrastructure*
2. *Automated transcription and enrichment services*
3. *Innovation and scalability of Transcribathon Platform*
4. *Dissemination and citizen science crowdsourcing campaigns*
5. *Project management and sustainability planning*

Svaka od tih aktivnosti sastoji se od nekoliko zadataka.

Content aggregation & data exchange Infrastructure obuhvaća agregaciju sadržaja u Europeanu, poboljšanje kvalitete podataka, razmjene podataka i standardizacije te validacije metapodataka. Sadržaj se dostavlja na Europeanu pomoću nacionalnih agregatora: organizacija koje prikupljaju, formatiraju i upravljaju metapodacima većeg broja pružatelja podataka, koriste vlastiti portal te služe kao pružatelji podataka za Europeanu. Nacionalni agregator za Europeanu u Hrvatskoj razvija se u sklopu projekta *e-Kultura – Digitalizacija kulturne baštine* koji predvodi Ministarstvo kulture i medija RH u suradnji s Hrvatskom radiotelevizijom, Nacionalnim i sveučilišnom knjižnicom u Zagrebu, Muzejom za umjetnost i obrt te Hrvatskim državnim arhivom. Agregator je trenutno u testnoj fazi u koju se uključio i Državni arhiv u Zagrebu.

Automated transcription and enrichment services obuhvaća razvoj HTR (*Handwritten Text Recognition*) tehnologije, transkripciju i prijevod metapodataka, otkrivanje tema i semantičko obogaćivanje.

Innovation and scalability of Transcribathon Platform sastoji se od razvoja i integracije HTR web plugina, standardizacije API-a Transcribathona, poboljšanja korisničkog iskustva unutar Transcribathon alata i ugrađivanja Transcribathona unutar Europeane CSP (*Core Service Platform*).

Dissemination and citizen science crowdsourcing campaigns sastoji se od organizacije niza događaja i aktivnosti koje imaju cilj doprijeti do javnosti i ustanova koje čuvaju kulturnu baštinu pomoću promoviranja Transcribathon događanja, organiziranja webinarima i radionica, diseminacije i uredničkih aktivnosti te procjene utjecaja projekta. Državni arhiv u Zagrebu će zajedno s Odsjekom za informacijske i komunikacijske znanosti Filozofskog fakulteta



u Zagrebu organizirati Transcribathon događaj na kojem će sudjelovati studenti različitih odsjeka Fakulteta.

Project Management and sustainability planning sastoji se od upravljanja projektom i financijama, sudjelovanja na PMB (*Project Management Board*) sastancima te omogućavanje dugoročne održivosti platforme.

Projekt će uključivati primjenu tehnologije obrade prirodnog jezika i velikih podataka za analizu transkripcija i njihovih prijevoda, pružajući podršku za semantičko obogaćivanje metapodataka, grupiranje i klasifikaciju. Istraživanja u kojima se koristi gradivo 19. stoljeća zahtijevaju angažman od nekoliko dana samo za dešifriranje rukopisa. Taj proces skratit će se na svega nekoliko minuta čime će istraživači uštedjeti dragocjeno vrijeme i omogućiti stvaranje još šire slike prošlih vremena.

Uloga Državnog arhiva u Zagrebu u projektu

Državni arhiv u Zagrebu aktivno sudjeluje u aktivnostima broj 1, 4 i 5. Glavni zadatak je dostavljanje novih materijala na Europeanu te je za u tu svrhu izabrano slijedeće gradivo:

- HR-DAZG-857 Zbirka Ulčnik Ivan, 1848./1930. – 1.650 komada
- HR-DAZG-839 Dragutin Hirc, 1844./1921. – 1.500 komada
- HR-DAZG-1 Poglavarstvo slobodnog kraljevskog grada Zagreba, Zapisnici sjednica Poglavarstva, 1800./1845. – 44 komada, 2.200 stranica
- Izbor pisama, 1832./1903. – 9 komada, 31 stranica
- HR-DAZG-24 Gradsko poglavarstvo Zagreb, Opskrbni odjel II-3, Potrošačke kartice, 120.621 komada

Potrošačke kartice ne pripadaju 19. stoljeću zbog čega je možda nejasno njihovo uključivanje u ovaj projekt. One su ranije već korištene kod zajedničkog projekta Odsjeka za informacijske i komunikacijske znanosti Filozofskog fakulteta u Zagrebu, tvrtke Bonsai Tech iz Zagreba te Državnog arhiva u Zagrebu, koji je započeo razvoj softvera za transkripciju strukturiranih tekstova kakve su potrošačke kartice. Zbog znatnog broja samih kartica i potrebe za velikom crowdsourcing platformom koja već ima široku i aktivnu publiku predloženo je i u konačnici prihvaćeno uključivanje potrošačkih kartica, a u konačnici i Državnog arhiva u Zagrebu, u projekt EnrichEuropeana+.

U osiguravanju novih sadržaja uz Državni arhiv u Zagrebu sudjeluju i The Provost, Fellows, Foundation Scholars and the other members of Board, of the College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin, Dublin City Council te Uniwersytet Wrocławski. Tako će se među novim materijalima na Europeanu moći pronaći pisma poljskog istraživača Stanisława Poniatowskog, pisma istraživača sanskrtu Alfreda Hillebrandta, zapisnici gradskih sjednica grada Dublina, pisma, dnevnici i bilježnice irskog pisca J. M. Syngea, pisma Oscara Wildea, spomenar Elsie Mahaffy, dnevnici Williama Hingstona i mnogi drugi zapisi.

Vrijednost projekta procijenjena je na 1.228.755 € (9.244.376,05 HRK), od čega CEF – Telecom financiranje sufinancira 75% odnosno 921.566 € (6.933.280,15 HRK), a ostalih 25% sufinanciraju partneri.

Državnom arhivu u Zagrebu ovo je prvi projekt vezan uz Europeanu, ali nadamo se da nije ni posljednji. ■

INFO

Više o EnrichEuropeana+ projektu saznajte na: <https://pro.europeana.eu/project/enricheuropeana>

Transcribathon platforma dostupna je na:

<https://www.transcribathon.eu/>

Više o Transkribusu saznajte na: <https://readcoop.eu/transkribus/>

