

Moderna
arhivistika

Časopis arhivske teorije in prakse
Journal of Archival Theory and Practice

Letnik 8 (2025), št. 2 / Year 8 (2025), No. 2

Moderna arhivistika

Časopis arhivske teorije in prakse

Journal of Archival Theory and Practice

ISSN 2591-0884 (online)

DOI: <https://doi.org/10.54356/MA>

Letnik 8 (2025), št. 2 / Year 8 (2025), No. 2

DOI: <https://doi.org/10.54356/MA/2025/2>

Izdaja / Published by:

Pokrajinski arhiv Maribor / Regional Archives Maribor

Glavna in odgovorna urednica / Chief and Responsible editor:

mag. Nina Gostenčnik, Pokrajinski arhiv Maribor, Glavni trg 7, SI-2000 Maribor,

telefon/ Phone: +386 2 228 5017; e-pošta/e-mail: nina.gostencnik@pokarh-mb.si

Uredniški odbor / editorial board:

- *mag. Nina Gostenčnik, Pokrajinski arhiv Maribor, Slovenija*
- *dr. Thomas Aigner, Diözesanarchiv St. Pölten, Avstrija*
- *dr. Bogdan Florin Popovici, Romunski državni arhiv, Romunija*
- *dr. Borut Batagelj, Zgodovinski arhiv Celje, Slovenija*
- *dr. Bojan Cvelfar, Mestna občina Celje, Slovenija*
- *mag. Nada Čibej, Pokrajinski arhiv Koper, Slovenija*
- *dr. Gregor Jenuš, Arhiv Republike Slovenije*
- *dr. Joachim Kemper, Institut für Stadtgeschichte Frankfurt am Main, Nemčija*
- *Leopold Mikec Avberšek, Pokrajinski arhiv Maribor, Slovenija*
- *dr. Miroslav Novak, Pokrajinski arhiv Maribor, Slovenija*
- *dr. Rik Opsommer, Stadsarchief Ieper - Universiteit Gent, Belgija*
- *Darko Rubčić, Državni arhiv u Zagrebu, Hrvaška*
- *dr. Izet Šabotić, Filozofski fakultet Univerziteta u Tuzli, Bosna in Hercegovina*
- *mag. Boštjan Zajšek, Pokrajinski arhiv Maribor, Slovenija*

Recenziranje / Peer review process:

Prispevki so recenzirani. Za objavo je potrebna pozitivna recenzija. Proces recenziranja je anonimen. / All articles for publication in the conference proceedings are peer-reviewed. A positive review is needed for publication. The review process is anonymous.

Lektoriranje / Proof-reading:

mag. Boštjan Zajšek, mag. Nina Gostenčnik

Prevajanje:

mag. Boštjan Zajšek (slovenščina), mag. Nina Gostenčnik (slovenščina, angleščina)

Oblikovanje in prelom / Design and typesetting: *mag. Nina Gostenčnik*

Objavljeni prispevki so prosto dostopni. Vse avtorske pravice ima izdajatelj Pokrajinski arhiv Maribor.

©Pokrajinski arhiv Maribor. Za prijavo in objavo prispevkov ni potrebno plačilo. / The publication offers open access to whole texts of the published articles. ©Pokrajinski arhiv Maribor. All articles are published free of charge.

Prejeto / Received: 11. 11. 2025

1.01 Izvirni znanstveni članek

1.01 Scientific article

<https://doi.org/10.54356/MA/2025/EETZ1766>

MOŽNOSTI UPORABE UMETNE INTELIGENCE V RODOSLOVJU: METODOLOŠKI PREMISLEK NA PRIMERU SLOVENSKE EMIGRANTSKE DRUŽINE LOŽAR/LOZAR IZ ELYJA V MINNESOTI

doc. dr. David HAZEMALI

Znanstveni sodelavec, Univerza v Mariboru, Filozofska fakulteta, Oddelek za zgodovino,
Maribor, Slovenija

david.hazemali@um.si

mag. Nina GOSTENČNIK

Arhivska svetovalka, direktorica, Pokrajinski arhiv Maribor, Slovenija

nina.gostencnik@pokarh-mb.si

asist. dr. Mladen BOROVIČ

asistent, raziskovalec, Univerza v Mariboru,
Fakulteta za elektrotehniko, računalništvo in informatiko, Slovenija

mladen.borovic@um.si

Izvleček:

Prispevek predstavlja metodološki premislek o možnostih uporabe umetne inteligence (UI) v rodoslovju na primeru slovenske emigrantske družine Ložar/Lozar iz Prigorice pri Ribnici, ki se ob koncu 19. in na začetku 20. stoletja preseli v Ely v Minnesoti. Izhodišče je Genealoški dokazni standard (GDS), ki določa »razumno izčrpno iskanje« virov, natančno navajanje, analizo in korelacijo informacij, razreševanje nasprotij ter oblikovanje preverljivega sklepa. Najprej pokažemo, kako klasična rodoslovna metoda, zasidrana v zgodovinske kritiki virov, vodi do dokaza o sorodstvenem razmerju med Frankom Lozarjem starejšim in Frankom Lozarjem mlajšim. Nato konceptualno obravnavamo tri skupine orodij UI – pogovorne sisteme, OCR/HTR in prepoznavanje imenovanih entitet (NER) – ter na študiji primera družine Ložar/Lozar pokažemo, kako lahko dopolnijo iskanje, transkripcijo in strukturiranje gradiva. Sklepno poudarimo, da morajo digitalni izhodi ostati hipoteze, ki dokazno veljavo pridobijo šele z verificiranjem po načelih GDS, in zagovarjamo hibridni model, v katerem UI deluje kot infrastrukturna podpora klasičnemu rodoslovnemu raziskovanju.

Ključne besede:

umetna inteligenca, digitalna humanistika, rodoslovje/genealogija, slovensko izseljenstvo, pogovorni sistemi, OCR, HTR, NER, Združene države Amerike

Abstract:

The Potential of Artificial Intelligence in Genealogy: A Methodological Reflection Using the Case Study of the Slovenian Emigrant Family Ložar/Lozar from Ely, Minnesota

This article offers a methodological reflection on the use of artificial intelligence in genealogy, based on the case study of the Slovenian emigrant family Ložar/Lozar from Prigoric near Ribnica, who settled in Ely, Minnesota, at the turn of the twentieth century. Our point of departure is the Genealogical Proof Standard (GPS), which requires reasonably exhaustive research,

accurate citation, analysis and correlation of information, resolution of conflicts, and a sound conclusion. We first show how the classical genealogical method, grounded in historical source criticism, leads to a documented proof of the kinship between Frank Lozar Sr and Frank Lozar Jr. We then conceptually examine three groups of AI-supported tools – conversational systems, OCR/HTR, and named entity recognition (NER) – and discuss how they can complement searching, transcription, and the structuring of sources. We conclude by arguing that digital outputs should be treated as hypotheses that acquire evidentiary force only through GPS-guided verification, and we advocate a hybrid model in which AI functions as infrastructural support to traditional genealogical research.

Key words:

artificial intelligence, digital humanities, genealogy, Slovenian emigration, chatbots, OCR, HTR, NER, United States

1 UVOD¹

Natančnost je temeljnega pomena za rodoslovne raziskave.

Brez nje bi bila družinska zgodovina fikcija.²

Rodoslovje (genealogija) kot pomožna zgodovinska veda ima dolgo tradicijo v evropski humanistiki. V zgodnjem obdobju je služilo predvsem političnim, cerkvenim in pravnim potrebam legitimiranja, v sodobnem raziskovalnem okolju pa se vse bolj uveljavlja kot metodološko zahtevna disciplina. Njegova epistemološka vrednost ni omejena na izdelavo družinskih dreves, temveč se kaže v sposobnosti natančne atribucije oseb, krajev in časovnih razmerij. S tem rodoslovje prispeva k širšemu zgodovinopisnemu razumevanju mobilnosti, lokalnih mrež in transnacionalnih stikov. Vloga rodoslovja je zato nepogrešljiva tako pri širšem zgodovinskem raziskovanju – kot uvodno atribucijsko orodje pri prepoznavanju oseb ali kot nosilec narativnih linij v biografskem, mikro- in lokalnem zgodovinopisju – kakor tudi v selitvenih študijah, kjer dopolnjuje demografske prikaze z mikroskopskimi, z dokazi podprtimi pripovedmi o poteh, odločitvah in družinskih strategijah.

V zadnjih dveh desetletjih je rodoslovje – podobno kot druge sorodne discipline in znanost kot taka – doživelo izrazito digitalno preobrazbo. Prvi korak je predstavljala obsežna digit(al)izacija arhivskih virov, ki je omogočila lažji dostop do razpršenih in institucionalno heterogenih korpusov. Naslednji, metodološko zahtevnejši korak pa pomeni vključevanje algoritmičnih pristopov, ki jih v javnem diskurzu običajno združujemo pod oznako umetna inteligenca (angl. *artificial intelligence*, AI; v nadaljevanju UI) in ki so vstopili v številne segmente znanstvenega dela (Chiarello et al., 2024; Deng in Lin, 2023; Ray, 2023; Dwivedi et al., 2023). V širšem znanstvenem polju raziskave UI pripisujejo dvojno vlogo: na eni strani deluje kot pospeševalec pri odkrivanju vzorcev v velikih podatkovnih zbirkah, na drugi pa odpira nova epistemološka in etična vprašanja glede razločljivosti modelov, kontrole nad podatki in porazdelitve tveganj (Gonzalez Garcia in Weilbach, 2023; Dwivedi et al., 2023). Ti razmisleki so neposredno

¹ Pričujoči prispevek je nastal v okviru podoktorskega temeljnega raziskovalnega projekta *Prilagoditev v stiski: južnoslovanski "sovražni tujci" iz Avstro-Ogrske v ZDA med prvo svetovno vojno* (evidenčna številka Z5-60179), ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije – ARIS.

² Izvirnik v angleščini: *Accuracy is fundamental to genealogical research. Without it, a family's history would be fiction* (Board for Certification of Genealogists, 2021, xix).

relevantni tudi za rodoslovje, saj ta že dolgo stoji na presečišču individualnih življenjskih zgodb, občutljivih osebnih podatkov in kvantitativnih populacijskih analiz.

V humanistiki, zlasti v zgodovinopisju, arhivistiki in dediščinski znanosti, se UI uveljavlja predvsem kot infrastruktura za delo z zgodovinskimi viri. Tehnike strojnega učenja se uporabljajo pri odkrivanju, opisu in klasifikaciji digitaliziranih objektov (Gîrbacia, 2024), avtomatska transkripcija rokopisov (angl. *Handwritten Text Recognition*, HTR) in optično prepoznavanje tiska (angl. *Optical Character Recognition*, OCR) pa v arhivih in knjižnicah postajata ključni tehnologiji za pretvorbo rokopisnega in tiskanega gradiva v strojno berljivo besedilo (Nockels et al., 2022; Ströbel et al., 2023; Vargas-Murillo et al., 2024). Vzporedno se razvijajo rešitve za prepoznavanje imenovanih entitet (angl. *Named Entity Recognition*, NER) in gradnjo podatkovnih modelov, ki povezujejo osebe, kraje in dogodke v kompleksna omrežja (Hendriks, Groth in van Erp, 2021). Ob tem pa študije opozarjajo tudi na vrzel med tehnološkimi možnostmi in dejansko pripravljenostjo uporabnikov (Jaillant, 2022; Jaillant in Aske, 2024).

Ko se ista infrastruktura prenese v raziskave, usmerjene v rekonstrukcijo življenjskih potekov, rodoslovnih dreves in migracijskih mrež, stopijo v ospredje vprašanja, ki so rodoslovju še posebej blizu. Študije kažejo, da lahko generativni modeli, utemeljeni na velikih jezikovnih modelih – VJM (angl. *large language models*, LLM) in arhitekture generiranja z razširjenim iskanjem (angl. *Retrieval-Augmented Generation*, RAG) učinkovito podprejo transkripcijo, prevajanje in obdelavo obsežnih korpusov zgodovinskega gradiva, vendar le ob jasno opredeljenih protokolih preverjanja in dokumentiranja rezultatov (Palladino, 2025; Drinkall et al., 2024). V širšem okviru demografskih in rodoslovnih študij so masovne digitalne baze (npr. milijonska rodoslovna drevesa, iz katerih so Kaplanis in sodelavci izračunali selitvene vzorce in stopnjo sorodstvenosti za 86 milijonov posameznikov) že pokazale, da lahko kombinacija rodoslovnih virov in naprednih izračunov radikalno razširi možnosti za analizo družinskih in selitvenih vzorcev (Kaplanis et al., 2018). Poskusi avtomatiziranega izluščanja rodoslovnih podatkov iz lokalnih družinskih kronik in množičnih evidenc kažejo, da so takšna orodja uporabna za hitro predlaganje kandidatnih ujemanj, a pri natančnosti in stabilnosti še vedno zaostajajo za specializiranimi, eksplicitno programiranimi pristopi (Stelter in Biehler, 2025). Sočasno nastajajo prispevki, ki z UI gradijo obsežne rodoslovne grafe, povezujejo razpršene zapise in odpirajo nove možnosti za analizo sorodstvenih mrež in migracijskih vzorcev (Wu et al., 2023), medtem ko razprave o forenzičnem rodoslovju in komercialnih bazah DNK opozarjajo na tveganja »umazanih« podatkov, zasebnosti in pravne nedorečenosti (Ramjee in Ringrose, 2020).

V rodoslovju ti razvojni trendi dobijo specifično obliko. Rodoslovno delo je po definiciji usmerjeno v povezovanje oseb, krajev in časovnih razponov čez več generacij, pri čemer se raziskovalec opira na razpršene, večjezične in pogosto nepopolne evidence. Z UI podprta digitalizacija in transkripcija omogočata, da so matične knjige, ladijski manifesti, popisi prebivalstva in časopisni zapisi dostopni kot iskalno in strojno obdelovalno besedilo; prepoznavanje imenovanih entitet olajša sledenje istim ali sorodnim akterjem čez različne vire; pogovorni sistemi pa lahko služijo kot vstopna točka za orientacijo po razpršenih korpusih, za generiranje variant imen in toponimov ter za preliminarno preverjanje skladnosti časovnic. Rodoslovcem se pri tem odpirajo obetavne možnosti, ki segajo onkraj zgolj izluščanja informacij – k interpretaciji dokumentov, kontekstualizaciji vsebin ter prepoznavanju jezikovnih in semantičnih vzorcev, a so za zdaj te zmožnosti v veliki meri še aspirativne, saj obstoječa orodja težko zanesljivo razlikujejo med resničnim in lažnim ter neprepričljivo argumentirajo svoje sklepe (Kansteiner, 2022; Hazemali et al., 2024). Raba UI v rodoslovju zato ne zmanjšuje,

temveč dodatno poudari klasična metodološka vprašanja: kako kakovostni in »kompletni« so uporabljeni podatki, kako razložljivi so postopki ter kako odgovorno ravnamo z občutljivimi osebnimi in družinskimi podatki (Ramjee in Ringrose, 2020).

V rodoslovju se tako posebej izrisujejo tri skupine tehnologij, ki neposredno posegajo v raziskovalni potek: (1) pogovorni sistemi oziroma chatboti, utemeljeni na VJM, ki delujejo kot interaktivni vmesnik za oblikovanje poizvedb, preliminarno sintezo in razlago konteksta; (2) tehnike digitalizacije in samodejne transkripcije, tj. OCR/HTR, ki pretvarjajo slikovne zapise v strojno berljivo besedilo; ter (3) obdelava naravnega jezika z osrednjim poudarkom na prepoznavanju imenovanih entitet (NER), ki iz besedila izlušči osebe, kraje, datume in druge ključne podatke ter jih pripravi za nadaljnjo analizo in povezovanje.

Pričujoči prispevek je zasnovan kot metodološki premislek o uporabnosti teh treh skupin orodij UI v rodoslovju, s posebnim poudarkom na raziskovanju slovenskega izseljevanja v Združene države Amerike ob koncu 19. in v začetku 20. stoletja. Empirično izhodišče predstavlja družina Ložar/Lozar iz vasi Prigorica pri Ribnici na Dolenjskem, katere člani so v tem času emigrirali v rudarsko naselje Ely na skrajnem severu zvezne države Minnesota. V središče postavljamo sorodstveno razmerje med Frankom Lozarjem starejšim (1836–1933) in Frankom Lozarjem mlajšim (1895–1918), ki sta vsak po svoje zaznamovala tamkajšnjo slovensko-ameriško skupnost.

Metodološko temelji izhodišče prispevka na Genealoškem dokaznem standardu – GDS (angl. *Genealogical Proof Standard*, GPS), ki skupaj s tradicionalnimi zgodovinopisnimi metodami dela predstavlja normativ za izčrpno iskanje, kritično analizo in preverljivo sklepanje v rodoslovju, ter na konceptualni kartografiji, s katero umeščamo orodja UI v posamezne faze raziskovalnega postopka. V prispevku ne merimo učinkovitosti algoritmov, temveč preizprašujemo pogoje njihove metodološke skladnosti na konkretni študiji primera: kje in pod katerimi pogoji lahko pogovorni sistemi, OCR/HTR in NER pospešijo pripravo gradiva, razširijo možnosti iskanja in izboljšajo sledljivost raziskovalnih odločitev – ob tem pa ostajajo sklepi strogo verificirani po načelih GDS.

Struktura članka sledi temu cilju. V metodološkem okviru (poglavje 2) najprej opredelimo odnos med GDS in uporabo orodij UI ter opišemo korpus virov in razloge za izbor študije. Nato v poglavju 3 rekonstruiramo klasični rodoslovni postopek na primeru družine Ložar/Lozar in iz njega izpeljemo nabor »kritičnih točk«, ki terjajo dopolnitev. V osrednjem četrtem poglavju predstavimo konceptualni prenos na orodja UI: od umeščenosti pogovornih sistemov v potek dela in protokolov njihove varne rabe, prek digitalizacije in transkripcije (OCR/HTR), do izluščanja in normalizacije entitet (NER). Peto poglavje je sklepno.

2 METODOLOŠKI OKVIR IN VIRI

2.1 Genealoški dokazni standard (GDS) in orodja UI

Metodološki izhodiščni okvir tega prispevka združuje genealoško dokazno prakso, kakor jo opredeljuje Genealoški dokazni standard (GDS), ter kritično tradicijo zgodovinopisne metodologije. GDS v literaturi običajno strnejo v pet medsebojno povezanih elementov: razumno izčrpno iskanje virov, natančno navajanje in vrednotenje uporabljenih virov, analizo in korelacijo informacij, razreševanje nasprotujočih si podatkov ter oblikovanje logičnega, preverljivega sklepa (Jones, 2013; Mills, 1997; Green, 2024). Zgodovinopisna notranja in zunanja kritika virov z GDS deli temeljno zahtevo po natančni atribuciji, razumevanju konteksta nastanka vira in triangulaciji z

neodvisnimi pričevanji; GDS temu dodaja izrazito normativno formulacijo in operativne standarde za rodoslovno sklepanje.

V tem okviru razumemo orodja UI kot pomoč raziskovalcu, ne njegovo nadomestilo: kot mehanizme za obvladovanje obsega, razpršenosti in jezikovne heterogenosti gradiva, ki pa jih je treba dosledno podvreči postopkom verifikacije in ponovljivosti. Ne gre torej za nadomestitev raziskovalčeve presoje z algoritmom, temveč za disciplinirano vključitev postopkov, ki skrajšajo pot od slike k besedilu, od besedila k entitetam, od entitet h kandidatnim ujemanjem in od hipoteze k preverjenemu dokazu. V poglavju 4 zato obravnavamo pogovorne sisteme, OCR/HTR in prepoznavanje ter normalizacijo entitet (NER) kot tri plasti istega cilja: omogočiti razumno izčrpnost iskanja v razpršenih korpusih, povečati transparentnost raziskovalnih korakov (sledljivost različic, poizvedb in odločitev) ter zmanjšati čas, ki ga klasična metoda sicer porabi za ročno prepisovanje in prečesavanje virov.

Skladno s tem vsak digitalni izhod obravnavamo kot hipotezo, ne kot dokaz: dokazno veljavo pridobi šele po triangulaciji z neodvisnimi primarnimi zapisi in vpetju v argumentacijo po načelih GDS. Posamezen računalniško podprt korak zato zahteva jasno revizijsko sled – zabeležene različice transkripcij, eksplicitno opisana pravila normalizacije, dokumentirane pragove in uteži pri ujemanjih ter časovno odtisnjene dnevniške poizvedb. Tak »audit trail« omogoča ponovljivost in zunanjo preverljivost raziskave ter etično odgovorno rabo UI v rodoslovnem kontekstu.

2.2 Študija primera: družina Ložar/Lozar

Empirično jedro prispevka predstavlja študija primera družine Ložar/Lozar iz Ribnice na Dolenjskem. Družina se je na prelomu iz 19. v 20. stoletje izselila v rudarsko naselje Ely v Minnesoti, ki je bilo v tem času eno ključnih središč slovenskega izseljenskega življenja na območju ameriške hribovske verige Iron Range (Klemenčič in Šeruga, 2019). Osrednja osebna figura v prvem delu študije je Frank (Franc) Lozar (rojen Ložar) starejši (1836–1933), rojen v Prigorici pri Ribnici, ki je v ZDA prispel že leta 1881 (Uredništvo, 1933) in se ustalil v Elyju, kamor je prišel prek sosednjega Towerja (Trunk, 1912, 497). Po pričevanjih sodobnikov je bil med prvimi slovenskimi priseljenci v tem še nastajajočem naselju in je postal eden vidnejših članov lokalne skupnosti (Peshel, 1933a). V svoji hiši – najverjetneje zgrajeni že leta 1888 – je gostil rojake, ki so prišli na delo v rudnike, prirejal verske obrede in postal gostitelj slovenskega duhovnika Frančiška Buha, ki je v Lozarjevi hiši opravil prva krsta in prve maše (Trunk, 1912, 497). V tej isti hiši je 15. avgusta 1889 nastalo Društvo sv. Cirila in Metoda, najstarejše slovensko podporno društvo v zvezni državi Minnesota, kar potrjuje, da je Lozarjeva vloga presegala raven zgolj zasebne gostoljubnosti in je segala v institucionalno organizacijo skupnosti (Zaplotnik, 1969, 74; Friš, 41). Čeprav Lozar ni zasedal vodilnih položajev v skupnosti, je veljal za njen tihi steber in za pionirja. Umrli je leta 1933 kot »najstarejši v Minnesoti živeči Slovenec«, obubožan in skoraj pozabljen (Peshel, 1933).



Slika 1: Hiša slovenskega pionirja Franka Lozarja v Elyju. Frank stoji drugi z desne (Trunk, 1912, 497).

Druga ključna osebnost je Frank (Franc, krščen Francišek) Lozar (rojen Ložar) mlajši (1894–1918), rojen 22. oktobra 1894 prav tako v Prigorici pri Ribnici (Krstna knjiga za obdobje 1886–1896). V Združene države je prišel kot otrok skupaj z očetom Lovrencem (bratom Franka starejšega) in bratom Janezom na ladji *SS La Bretagne* leta 1900 (Manifest ladje S.S. La Bretagne, 1900), medtem ko je mama Neža (rojena Bartol) prispela z ladjo *SS La Touraine* dve leti kasneje (Passenger and crew lists of vessels arriving at New York, 1897–1942). Iz popisa prebivalstva iz leta 1905 je razvidno, da je družina Lozar (priimek je postal amerikaniziran) takrat že živela v Elyju, kjer je postala bogatejša še za enega sina – Antona (Lozar, Frank, *Family Search*). Oče je umrl kmalu po prihodu, zato je Frank že v mladih letih prevzel finančno breme družine. Bil je podjeten, odprl je manjšo trgovino in se v Elyju uveljavil kot dejaven in spoštovan član skupnosti (Poročevalec, 1921). Do ameriškega vstopa v prvo svetovno vojno je Frank uspešno pridobil ameriško državljanstvo (Pogorelec, 1921). Leta 1917 je kot ameriški državljan, tako kot številni njegovi rojaki, vstopil v oborožene sile ZDA in leto pozneje padel v bojih pri St. Mihielu v Franciji (Hazemali, 2023, 496; Van Brunt, 1921, 645). Lozar se je zapisal v anale kot prvi Elyjčan, ki je padel v boju, in kot edini ugotovljeni, ki je posthumno prejel spominsko diplomu, v kateri ga je poveljnik ameriških ekspedicijskih sil John J. Pershing proglasil za vojnega heroja (Spominska diploma na Franka Lozarja). Elyjska izpostava veteranske organizacije American Legion je po prvi svetovni vojni izpostavo poimenovala po njem – Frank Lozar Post, kar je pomenilo izjemno čast (Uredništvo, 1921).

Rodoslovna raziskava, ki je vključevala tudi terensko delo na tamkajšnjem elyjskem pokopališču, je pokazala, da sta bila Frank Lozar starejši in Frank Lozar mlajši v neposrednem sorodu: pionir Frank Lozar je bil brat Lovrenca Lozarja, očeta vojnega heroja, in Frankov stric. Ložarjevi oziroma Lozarjevi so torej pustili neizbrisljiv pečat na Elyju, tamkajšnji slovensko-ameriški skupnosti pa tudi na ameriških oboroženih silah, ki so bile vpoklicane z območja Iron Rangea.



Slika 2: Fotografija slovensko-ameriškega vojnega heroja iz Elyja Franka (Frančiška) Lozarja (rojenega Ložar) v vojaški uniformi, najbrž leta 1918³

Študija primera Ložar/Lozar je bila izbrana, ker v zgostitvi prikaže tri sistemske težave, značilne za slovensko, nemškoavstrijsko in ameriško izseljensko gradivo konca 19. in začetka 20. stoletja. Prva je večjezičnost in paleografija: v evropskem segmentu prevladujejo slovenski, nemški in latinski zapisi, pogosto v rokopisnih pisavah (npr. v nemški kurzivi *Kurrentschrift*⁴), ki otežujejo natančno branje in transkripcijo; ameriški segment je praviloma v angleščini, a z močno prisotnimi fonetičnimi transliteracijami slovenskih imen. Druga je ortografska variabilnost: isti akterji in kraji se pojavljajo v različnih, pogosto fonetičnih ali jezikovno posredovanih oblikah (Ložar/Lozar; Ribnica/Ritnica/Rahnica/Reinfnica/Reifnitz)⁵, pri čemer diakritični znaki pogosto

³ Fotografijo in Lozarjevo vojaško uniformo hrani Zgodovinski muzej Ely-Winton (Ely-Winton History Museum).

⁴ Kurrent je stara oblika nemškega rokopisa, ki temelji na poznosrednjeveškem kurzivu, znanem tudi kot *Kurrentschrift*, *deutsche Schrift* in nemški kurziv.

⁵ Popačena imena Lozarjevega rojstnega kraja so nakazovala tako na nemški toponim Reifnitz, ki v tem našem primeru pomeni slovenski kraj Ribnico na Dolenjskem, kot tudi na Ribnico ob Vrbskem jezeru

odpadejo ali se prevedejo v približke. Tretja je arhivska razpršenost in nezdružljivost popisnih shem: matične knjige iz obdobja Avstro-Ogrske, ladijski manifesti, ameriški popisi prebivalstva, vloge za pridobitev ameriškega državljanstva, vojaški dokumenti, lokalno časopisje in pokopališke evidence nastajajo v različnih institucionalnih okoljih, z različnimi minimalnimi nabori podatkov in brez enoličnih identifikatorjev. V takem okolju klasična metoda sicer ostaja uporabna, a je časovno zahtevna in izrazito odvisna od izkušenj raziskovalca pri sklepanju iz posrednih dokazov.

2.3 Korpus virov: značilnosti, dostopnost in kakovost

Rodoslovna analiza primera Ložar/Lozar temelji na kombinaciji slovenskih, nemškoavstrijskih in ameriških primarnih in sekundarnih virov, ki skupaj tvorijo raznolik, večjezičen korpus. Na evropski strani so osrednji vir cerkvene matične knjige župnije Ribnica (krstne in rojstne knjige, poročne ter mrliške knjige), ki zajemajo prebivalce vasi Prigorica in širše okolice. Te knjige, deloma že digitizirane⁶, se razlikujejo po berljivosti in jeziku: v obravnavanem časovnem intervalu pogosto prehajajo med slovenščino in nemščino, mestoma tudi latinščino, zapisane so v različnih pisavah, kar od raziskovalca zahteva paleografsko znanje in poznavanje lokalnih konvencij. V kontekstu rodoslovja predstavljajo matične knjige najzanesljivejšo osnovo za vzpostavljanje sorodstvenih vezi, saj praviloma natančno beležijo razmerja med starši, otroki in botri, obenem pa prinašajo nevarnosti napačnih branj, izgubljenih vpisov in jezikovno posredovanih interpretacij, ki jih je treba obravnavati z metodami klasične notranje in zunanje kritike vira.

Prehod čez Atlantik osvetljujejo ladijski manifesti, denimo za čezoceanki *SS La Bretagne* in *SS La Touraine*, na katerih so bili vpisani posamezni člani družine Ložar/Lozar. Ti dokumenti so ključni za določitev časovnice in smeri migracij (kraj vkrcanja, približen čas poti, deklarirani kraj izvora), vendar njihova uporabnost ni brez omejitev: zapisi so pogosto fonetično prilagojeni, imena popačena, kraji izvorov napačno transkribirani, podatki o starosti, poklicu in družinski strukturi pa – če sploh – podani v minimalnem obsegu. Zato imajo ladijski manifesti v raziskovalnem postopku predvsem orientacijsko vlogo: omogočajo zasnovo delovne hipoteze o selitveni poti, ki jo je treba preverjati in dopolnjevati z drugimi viri.

Ameriški segment virov sestavljajo zvezni in državni popisi prebivalstva (zlasti popis prebivalstva na zvezni ravni iz let 1910 in 1920 ter popis zvezne države Minnesote iz leta 1905), evidence naturalizacij posameznikov, vojaški naborni registri in žrtveniki, obsežno časopisje ter pokopališke evidence, dopolnjene s terenskim delom na elyjskem pokopališču. Popisi prebivalstva so dragoceni za rekonstrukcijo celotnih gospodinjstev in njihovo umeščanje v prostor (naslovi, poklici, sobivanje sorodnikov), a trpijo zaradi napak pri navedbi starosti, letnice prihoda in (pogosto) napačno navedenega kraja rojstva. Časopisni viri prinašajo biografske drobce in vpogled v socialni ter kulturni kontekst skupnosti, obenem pa nosijo tveganje subjektivnosti, lokalnih pristranskosti in homonimij. Pokopališke knjige in terenske meritve grobov omogočajo verifikacijo datumov smrti in lokacij pokopa, s čimer zaokrožijo življenjski lok posameznikov, a so omejene na zadnjo fazo biografije. Vse našteje skupine virov so zato uporabljene

(nem. *Reifnitz am Wörthersee*) na avstrijskem Koroškem. Takšne in podobne toponimske zagate so stalnice v rodoslovju.

⁶ Matične knjige župnij Nadškofij Ljubljana in Maribor ter Škofije Koper so pred nekaj leti postale dostopne na platformi *Matricula Online*: <https://data.matricula-online.eu/sl/slovenia/>; pokritost se med župnijami razlikuje (mestoma posamezne knjige manjkajo).

komplementarno: njihove prednosti izkoriščamo v korelaciji, njihove slabosti pa upoštevamo pri vrednotenju dokazne teže posamezne informacije.

3 KLASIČNA RODOSLOVNA METODA NA PRIMERU DRUŽINE LOŽAR/LOZAR

V okviru klasične rodoslovne prakse, ki je neodtujljivo vpeta v zgodovinopisno kritiko virov, Genealoški dokazni standard (GDS) ne deluje kot togo zaporedje predpisov, temveč kot normativni horizont, znotraj katerega se konkretno delo vedno znova prilagaja vrsti in kakovosti virov, (ne)dostopnosti korpusov, institucionalnim omejitvam ter individualnemu sosledju raziskovalčevih korakov in njegovim kontekstualnim interpretacijam. Na podlagi metodološkega okvira in pregleda virov, predstavljenih v poglavju 2, v tem poglavju korak za korakom rekonstruiramo klasični rodoslovni postopek na primeru družine Ložar/Lozar: od izhodiščnega »razumno izčrpnega iskanja« čez vzpostavljanje družinske konfiguracije, preverjanje migracijskih poti in razreševanje nasprotij med viri, do končnega sklepa o sorodstvenem razmerju med Frankom starejšim in Frankom mlajšim. V središču so zmožnost razlikovanja med virom, informacijo in dokazom, tehtanje verodostojnosti ter prepričljivo koreliranje razpršenih podatkov. Študija primera Ložar/Lozar tako deluje dvojno: kot konkretna demonstracija, kako GDS vodi do sklepa, ki prestane preizkus ponovljivosti in notranje koherence, in hkrati kot matrica, na kateri bomo v naslednjem poglavju natančno pokazali, kje bi lahko posamezne korake klasične metode smiselno dopolnili z orodji UI.

Prvi korak je obsegal izčrpno iskanje virov in literature po načelu »od bližnjega k oddaljenemu« in hkrati »od zanesljivejšega k negotovemu«. Izhodišče so bile cerkvene matične knjige župnije Ribnica (krsti, poroke, smrti) za vasi v okviru župnije, s posebnim poudarkom na Prigorici, v katerih se je v časovnem loku 1830–1900 izrisala družinska konfiguracija z dvema ključnima akterjema: Frankom (rojen Franc) Ložarjem starejšim (1836–1933) in njegovim bratom Lovrencem (1844–1907), očetom Franka mlajšega (1894–1918). Šele na tej osnovi, z naborom t. i. stabilnih identifikatorjev (osebna imena, starši, krajevni in časovni označevalci ipd.), je bilo smiselno poseči v heterogenejše korpuske: ladijske manifeste, ameriške popise prebivalstva, naturalizacijske evidence, vojaško gradivo, pokopališke knjige in lokalno časopisje. Izhodiščna hipoteza o selitveni poti je dobila prvi okvir v ladijskih seznamih: za »pionirsko« generacijo je ključno, da je Frank Ložar starejši v ZDA priplul že leta 1881 na ladji *SS Alaska* in se iz bližnjega Towerja najverjetneje že leta 1888 »s sekiro v roki« podal v nastajajoči Ely. Do 29. marca 1889 je v mestu stala njegova prostorna hiša, v kateri je duhovnik Frančišek Buh krstil »prvo Slovenko«, 15. avgusta istega leta pa so slovenski rudarji iz Elyja in Towerja v tej hiši ustanovili najstarejše slovensko-ameriško bratsko-podporno društvo v Minnesoti – Društvo sv. Cirila in Metoda. Navedeni datumi in dogodki se ujemajo v več neodvisnih zapisih (časopisnih noticah in nekrologih, spominskih pričevanjih sodobnikov in poznejših krajevnih zgodovinopisnih sintezah) ter v kombinaciji s pokopališkimi evidencami in terenskim delom *in situ* (Frank starejši umre leta 1933 kot »*najstarejši v Minnesoti živeči Slovenec*«) tvorijo trden lok, ki Franka Lozarja starejšega umesti v središče institucionalne in dušnopastirske organizacije slovensko-ameriške skupnosti v Elyju.

Druga, mlajša migracijska veja – družina Lovrenca Ložarja – je bila na temelju izrisane družinske konfiguracije iz matičnih knjig rekonstruirana z dvojico ladijskih prihodov in vmesnimi popisi. Oče z sinovoma Frankom in Janezom (pozneje Johnom) prispe leta 1900 na *SS La Bretagne*, mati Neža (roj. Bartol) pa leta 1902 na *SS La Touraine*. Popis Minnesote leta 1905 potrdi družino v Elyju (z novorojenim Antonom) in

Franka mlajšega označi kot »učenca«, kar je dragocen starostni in družbeni podatek, ki se prepričljivo ujema z nadaljnjo družinsko časovnico: smrt očeta leta 1907, Frankovo šolanje, odprtje manjše trgovine z mešanim blagom, ureditev državljanstva v Duluthu pred vstopom ZDA v prvo svetovno vojno, prostovoljna pridružitve oboroženim silam, bojevanje in smrt v Franciji leta 1918 ter odmevni pogreb v domačem Elyju z novembra 1919 ustanovljeno izpostavo American Legion, poimenovano po njem. Vse našeto spremlja dosledno beleženje arhivov, vrst dokumentov, datacij, signatur oziroma trajnih povezav – ne kot formalizem, temveč kot pogoj, da je kasnejša korelacija preverljiva.

Jedro metodološke teže v tem primeru – in s tem prikaz prednosti ter meja klasične metode – leži v sprotne, »ročnem« sistematičnem ocenjevanju verodostojnosti in razreševanju nasprotij, ki izvirajo iz večjezičnosti, paleografske zahtevnosti in ortografske variabilnosti. Cerkevne matične knjige ribniške župnije so, kot velewa praksa, zanesljiv vir za rojstne podatke in sorodstvena razmerja, a terjajo paleografsko kompetenco (npr. branje vpisov v *Kurrentschriftu* in izmenjavo slovenščine, nemščine ter latinščine). Nasprotno so ladijski manifesti in ameriški popisi obremenjeni s fonetičnimi zapisi imen in toponimov ter nekonsistentnostjo pri starostih in letnicah priselitve, časopisje in spominska literatura pa sta neprecenljiva za rekonstrukcijo družbene vloge posameznikov, vendar vpelujeta homonimije in selekcijske pristranskosti. Na mikro ravni to vidimo pri Franku mlajšem v nizu odklonov pri rojstnem kraju: »Ribnica«, »Ritnica«, »Rahnica«, »Reinfnica« in celo »Reifnitz« – ob variacijah letnice rojstva 1894/1895 –, ki se pojavljajo v različnih dokumentih. GDS na tej točki zahteva vrnitev k primarnemu, časovno najbližjemu zapisu: krstna knjiga župnije Ribnica izpričuje rojstvo Franka Lozarja kot Frančiška Ložarja 22. oktobra 1894 v Prigorici (oče Lovrenc, mati Neža, rojena Bartol; krst naslednjega dne kot Frančišek). Fonetične in jezikovne variante se s tem prevedejo v verodostojno normalizacijo: »Ritnica/Rahnica/Reinfnica/Reifnitz« so v danem kontekstu fonetični približki za Ribnico (nem. *Reifnitz*) na Dolenjskem in ne koroškega kraja Reifnitz ob Vrbskem jezeru. Ta korekcija ni zgolj jezikoslovna vaja, temveč odločilen korak, ki omogoči ujemanje Franka mlajšega čez ladijske sezname, popise, naturalizacijske in vojaške registre ter časopisne omembe; veriga »rojstvo–migracija–naturalizacija–služenje–smrt–spomin« se zaključi v podatku o padcu 13. septembra 1918 v Franciji in v poznejši spominski praksi.

Najobčutljivejši del klasičnega rodoslovnega postopka v tej študiji primera je bilo razločevanje dveh soimenjakov in razreševanje protislovij, ki iz tega neizogibno izhajajo. Frank starejši, pionir in posestnik, je v krajevni tradiciji in zgodovinopisnih sintezah dokumentiran kot soustanovitelj institucionalnega življenja Slovencev v Elyju, medtem ko se je Frank mlajši v krajevno in ameriško vojaško zgodovino vpisal kot proglašeni vojni heroj slovenskih korenin. Ker se oba pojavljata v fragmentiranih, mestoma nejasnih noticah in zabeležbah v različnih virih ter celo na marmornih ploščah v šoli in na elyjskem pokopališču, je bilo treba uvesti strogo hierarhijo virov: (a) za rojstne podatke in sorodstvena razmerja so bile odločilne matične knjige; (b) za migracijsko časovnico ladijski manifesti in ameriški popisi; (c) za služenje in smrt vojaški ter veteranski zapisi (vključno z *Honor Roll*); (č) za trajno prostorsko korelacijo in končno verifikacijo pokopališke knjige in terenska verifikacija. Prav pokopališki odsek 16 s parcelama 43 (Frank mlajši) in 48 (Frank starejši) je deloval kot materialni korekcijski element, ki je omogočil dokončno razmejitev in povezavo: Frank starejši je brat Lovrenca, torej stric Franka mlajšega. S tem je bil izpolnjen četrti pogoj GPS (razreševanje nasprotij) in ustvarjena podlaga za sklep, ki ni le verjeten, marveč dokazno podprt v več neodvisnih, med seboj konsistentnih zapisih.

Končni sklep, v duhu petega pogoja GDS, zato ni zgolj deklarativen, temveč izhaja iz verige medsebojno potrjujočih se evidenc: (1) Frank (Franc) Lozar starejši (1836–

1933) je v Elyju ob koncu 19. stoletja bistveno prispeval k vzpostavitvi materialnega in organizacijskega jedra slovensko-ameriške skupnosti; (2) Frank Lozar mlajši (1894–1918) je kot naturalizirani državljani ZDA padel v boju za svojo »novo domovino« 13. septembra 1918 in bil proglašen za vojnega heroja; (3) oba sta pokopana v isti sekciji elyjskega pokopališča, kar pomeni ne le družinsko bližino, temveč – v spominski krajini – tudi kontinuiteto priseljenske skupnosti.

Čeprav v tem poglavju kažemo na visoko razločevalno moč klasične metodologije, istočasno razgaljamo tudi njene omejitve: časovna zahtevnost ročnega branja in prepisovanja, večjezičnost in variabilnost pravopisja (Ložar/Lozar; Ribnica/Ritnica/Rahnica/Reinfnica/Reifnitz), neenotni arhivski standardi, poškodovanost oziroma neberljivost gradiva in odsotnost enoličnih identifikatorjev v virih. Prav na teh mestih se odpre prostor za dopolnitev z digitalnimi pristopi, podprtimi z UI: pogovorni sistemi kot izhodišče, ki z večjezičnimi poizvedbami in povzemanjem olajšajo orientacijo po razpršenih korpusih; avtomatska transkripcija (HTR) tam, kjer pisava ovira iskanje; in izluščanje ter normalizacija entitet (NER) tam, kjer variabilnost imen zabriše sled – vselej ob pogojih preverljivosti in sledljivosti, ki jih predpisuje GPS. V naslednjem poglavju zato konceptualno prikažemo, kako bi lahko z UI podprti postopki smiselno dopolnili vsakega od opisanih korakov klasične analize.

4 ORODJA UI V RODOSLOVNEM DELOVNEM TOKU

4.1 Pogovorni sistemi (chatboti)

V zadnjih letih so pogovorni sistemi, utemeljeni na VJM, postali vsakdanji del raziskovalnega okolja tudi v humanistiki. Konceptualne in empirične študije jih obravnavajo kot »inteligentne pomočnike«, ki lahko podprejo rutinske, a časovno potratne naloge – od povzemanja in preoblikovanja besedil, prevajanja in urejanja citatov do osnovnega kodiranja in oblikovanja iskalnih poizvedb (Kasneci et al., 2023; Deng in Lin, 2023; Carretero in Gartner, 2024; Khudair, 2025; Palladino, 2025). Raziskave s področja izobraževanja in akademskega pisanja opozarjajo, da so takšna orodja retorično in slogovno prepričljiva, vendar pogosto mešajo točne in netočne podatke, »halucinirajo« ter reproducirajo obstoječe pristranskosti (Ray, 2023; Currie, 2023; Stokel-Walker, 2023; Else, 2023; Altmaë, Sola-Leyva in Salumets, 2023; Buholayka, Zouabi in Tadinada, 2023; Tirado-Olivares et al., 2023). V tej luči se utrjuje normativ, da so pogovorni sistemi dopustni kot podpora raziskovalcu, ne kot nadomestilo njegove presoje: potek dela mora ostati transparenten, vse trditve, ki vstopijo v znanstveno argumentacijo, pa preverljive v primarnih virih (Lozić in Štular, 2023; Hazemali et al., 2024).

V arhivistiki in zgodovinopisju je razprava o pogovornih sistemih še posebej intenzivna, ker njihov »človeški« način odgovarjanja zlahka ustvari iluzijo razumevanja in gotovosti. Hazemali in sodelavci so pri testiranju pogovornega sistema, ki ga poganja VJM GPT-3.5, nad historičnim gradivom pokazali, da je ta uporaben pri lokaliziranih nalogah – izluščanje imen, datumov, osnovnega vsebinskega okvira –, skoraj povsem nezanesljiv pa pri interpretativnih vprašanjih, sklepih o vzročnosti in vrednotenju virov (Hazemali et al., 2024). Podobno Carretero in Gartner (2024) opozarjata, da generativna orodja olajšajo dostop do zgodovinskih virov, a lahko z retorično samozavestjo svojih odgovorov zakrijejo razlike med virom, povzetkom in interpretacijo. Ključno metodološko vprašanje zato ni, ali pogovorne sisteme uporabljati, temveč kako jih vpeljati v raziskovalni potek, ne da bi pri tem razrahljali standarde dokazovanja in citiranja.

V rodoslovju se ta vprašanja še zaostrijo, saj se raziskovalec giblje na presečišču občutljivih osebnih podatkov, večgeneracijskih sorodstvenih mrež in pogosto fragmentarnih virov. Študije iz uvoda potrjujejo, da pogovorni sistemi in sorodna orodja lahko avtomatizirajo del postopkov izluščanja, prevajanja in preliminarne analize besedil, pomagajo pri odkrivanju vzorcev v velikih rodoslovnih zbirkah (Drinkall et al., 2024) in demokratizirajo dostop do naprednejših analitičnih tehnik. Hkrati pa ostajajo nerešeni problemi kakovosti podatkov, zasebnosti in etike, zaradi česar avtorji generativno UI vidijo predvsem kot orodje za predlaganje kandidatnih ujemanj in orientacijo, ne pa kot »ekstraktor resnice« (Stelter in Biehler, 2025; Hazemali et al., 2024; Jaillant in Aske, 2024).

Na tej podlagi obravnavamo v nadaljevanju pogovorne sisteme kot dopolnilo klasične rodoslovne metode pri nizu konkretnih mikronalog v študiji družine Ložar/Lozar – od oblikovanja večjezičnih poizvedb in generiranja variant imen do orientacije po institucionalno razpršenih korpusih in preliminarne preverjanja skladnosti časovnic.

Prva je iztočno iskanje ali »googlanje« v naravnem jeziku, kjer pogovorni sistem z dostopom do spleta prevzame vlogo posrednika med raziskovalcem in razpršenimi spletnimi korpusi. Namesto zaporedja ročnih poizvedb v klasičnem iskalniku (npr. *Lozar Ribnica site:familysearch.org*; *Lozar Ely Minnesota*) lahko raziskovalec zastavi širše, opisno vprašanje: »V katerih spletnih rodoslovnih zbirkah, forumih ali digitalnih arhivih se pojavlja priimek Ložar/Lozar v povezavi z Ribnico na Dolenjskem?« Odgovor – če pogovorni sistem dela nad dejanskimi viri in jih tudi citira – omogoča hitro prepoznavo potencialno relevantnih portalov (komercialne baze, forumi, krajevni projekti digitalizacije), ki jih nato raziskovalec klasično preveri.

Drugi sklop nalog se nanaša na generiranje in preverjanje variant osebnih imen in toponimov, kar je v primeru družine Ložar/Lozar ključnega pomena. Poziv v slogu: »Naštev verjetne zgodovinske in fonetične variante priimka 'Ložar' in kraja 'Ribnica' v nemških, angleških in latinskih zapisih, posebej okoli 1900«, lahko v nekaj sekundah vrne nabor možnih oblik (npr. Lozar, Loschar, Lotzar; Ribnitz, Riebntz) skupaj z opozorilom na nevarne bližnjice (npr. Reifnitz na Koroškem, ki ga v našem primeru ne smemo zamenjati z Ribnico na Dolenjskem). V praksi so prav takšne variantne oblike (»Ritnica«, »Rahnica«, »Reinfnica« in »Reifnitz«) igrale ključno vlogo pri razreševanju nasprotujočih si navedb v ameriških virih in njihovi normalizaciji na Ribnico – kar pa je bilo mogoče šele po vrnitvi h krstnemu vpisu v ribniški matični knjigi.

Tretja skupina nalog je povezana z orientacijo po razpršenih institucionalnih okoljih in oblikovanjem raziskovalnega načrta. Pogovorni sistem je tu uporaben kot nekakšen »digitalni referent«, ki odgovarja na vprašanja tipa: »Kje so dostopne digitizirane matične knjige za župnijo Ribnica?«, »Kateri ameriški popisi pokrivajo mesto Ely v Minnesoti okoli let 1900 in 1910?« ali »V katerih zbirkah so objavljeni ladijski manifesti za ladji SS La Bretagne in SS La Touraine?« Če orodje korektno navaja povezave na arhive in zbirke, lahko raziskovalec v nekaj minutah sestavi delovni seznam repozitorijev in zbirk – nalogo, ki je bila v preteklosti pogosto časovno precej zahtevna.

Četrty tip nalog je podpora pri interpretaciji posameznih dokumentov in preverjanju notranje logike podatkov. Rodoslovec se lahko pri nejasnem zapisu obrne na pogovorni sistem z vprašanjem: »Prepiši in prevedi naslednji latinski krstni vpis iz leta 1894 ter razloži osnovno strukturo zapisa«, ali pa: »Ali se ti zdijo naslednje letnice rojstev staršev in otrok v družini Ložar med 1860 in 1900 notranje skladne?«. Izkušnje kažejo, da so VJM pri razlagi standardiziranih obrazcev in starih jezikovnih formul pogosto zelo uporabni, pri preverjanju časovne skladnosti pa lahko hitro opozorijo na očitne nelogičnosti (npr. premlada mati, otrok, rojen po očetovi smrti).

Peti sklop predstavljajo podporne, »paragenealoške« naloge, ki sicer ne prinašajo novih dejstev, a razbremenijo raziskovalca: povzemanje obsežnih člankov o emigraciji v Minnesoti, oblikovanje prvih osnutkov sorodstvenih sestavkov (npr. o Franku Lozaru mlajšem), generiranje seznamov iskalnih besed za digitizirane časopise (kombinacije imen in priimkov, krajev, ladij) ali sprotno zapisovanje raziskovalnega dnevnika: »Na kratko povzemite naslednje zapiske kot kronološki potek raziskave«. V vseh teh primerih UI ne nastopa kot »avtor« rodoslovne študije, temveč kot orodje za obvladovanje obsega in za organizacijo dela.

Vse našteje naloge kažejo, da je mestu pogovornih sistemov v rodoslovju treba odmeriti jasno metodološko vlogo: delujejo na ravni predlogov, normalizacij in organizacije, ne pa na ravni dokazov. V skladu z GDS vsako trditev, ki vstopi v rodoslovno argumentacijo, utemeljujemo z natančno citiranim, preverljivim virom; pogovorni sistem nam lahko pomaga, da do vira hitreje pridemo, da variacije imen prej prepoznamo in da preliminarno preverimo skladnost časovnic, nikoli pa sam po sebi ne zadostuje kot dokaz.

4.2 Optično prepoznavanje znakov in strojno prepoznavanje rokopisov (OCR/HTR)

V klasičnem rodoslovnem raziskovanju je ovira pogosto že prvi korak: kako sploh priti do berljivega besedila. Matične knjige, zemljiške knjige, lokalne kronike, pisna korespondenca in drugo arhivsko gradivo praviloma obstajajo kot tiskani ali rokopisni dokumenti, pogosto v starejših pisavah in v več jezikih. Optično prepoznavanje znakov (OCR) in strojno prepoznavanje rokopisov (HTR) zato postajata osrednja tehnološka »vrata« za rokovanje s temi dokumenti. V zadnjem desetletju je področje doživelo izrazit prehod od klasičnih in eksperimentalnih metod k pristopom globokega učenja, npr. s hibridnim modelom tipa CNN-LSTM-CTC in transformerjem, kar je bistveno izboljšalo natančnost tudi pri zgodovinskih in rokopisnih besedilih (Bluche in Messina, 2017; Liang et al., 2020; Khan et al., 2025; Liu et al., 2025).⁷

Sistematične evalvacije kažejo, da dobro izurjeni modeli pri standardiziranih zgodovinskih tiskanih pisavah (npr. črnotiskarske knjige in časniki) dosegajo napake na ravni znaka (CER) okrog 1–2 %, v najboljših primerih tudi pod 1 % (Ströbel, Clematide in Volk, 2020; Muehlberger, 2019). Za enotnejše rokopisne korpuse (npr. administrativne knjige, serijske rokopisne zbirke) so v literaturi prikazani rezultati z napako na ravni znaka (angl. *character error rate*, CER) pod 5 % – izjemoma okrog 2 % –, medtem ko se pri bolj heterogenih ali grafično zahtevnih rokopisih napaka pogosteje giblje med približno 10 in 20 % (Muehlberger, 2019; Ingle et al., 2019). Empirični prag, ki ga navaja Transkribus/READ, je okoli 10 % CER: pod to mejo so avtomatski prepisi praviloma uporabni tudi kot osnova za branje in nadaljnje obdelave, nad njo pa se njihova vrednost premakne predvsem v smeri iskanja in indeksiranja, ki ostajata učinkovita tudi pri precej višjih napakah (Muehlberger, 2019; Ströbel, Clematide in Volk, 2020; Ingle et al., 2019; Neto et al., 2024). Ključni dejavniki so kakovost digitalnih posnetkov (ločljivost, kontrast, poškodovanost papirja), homogenost pisave in obseg ter kakovost »ground truth« prepisov, na katerih se modeli učijo. V rodoslovju in zgodovinski demografiji to pomeni, da so najprimernejši kandidati za HTR serijski, relativno enotno vodeni registri, kot so

⁷ Za ponazoritev tehnološkega prehoda: razvoj OCR/HTR modelov je kot prehod od ročnega sestavljanja ure (klasične metode), ki je zahtevalo natančno prilagoditev vsakega zobnika (ročna ekstrakcija značilnosti), do pametne ure (globoko učenje), ki se nauči sama prilagajati in zaznavati vzorce (CNN, LSTM), pri čemer so transformatorji kot dodatek GPS in umetne inteligence, ki omogočajo globalno in kontekstualno razumevanje v realnem času, ne glede na kompleksnost podlage.

matične knjige, registri prebivalstva, serije tesno sorodnih obrazcev itn., medtem ko fragmentarni, močno poškodovani ali grafično zelo heterogeni viri še naprej zahtevajo večji delež ročnega dela.

HTR se je v zadnjem desetletju vedno bolj uveljavil kot pomemben del digitalizacijskega delovnega toka v arhivih in knjižnicah, zlasti na platformi Transkribus, ki omogoča učenje prilagojenih modelov na osnovi ročno prepisanih vzorcev in nato serijsko transkripcijo večjih korpusov (Muehlberger et al., 2019; Nockels et al., 2022). Nockels s sodelavci v sistematičnem pregledu kaže, da večina objav, ki obravnavajo Transkribus, izhaja iz humanistike in arhivskih ter knjižničnih okolij, pogosto v zvezi z rokopisnimi knjigami in registri, pri čemer se del projektov eksplicitno navezuje na rodoslovna in lokalno-zgodovinska vprašanja (Nockels et al., 2022). Podobno druge študije poudarjajo, da je prehod od digitalne slike k strojno berljivemu besedilu danes osnovni pogoj za polnobesedilno iskanje (angl. *full-text search*), kvantitativne analize in avtomatizirano povezovanje zapisov (Miyagawa, 2023; Lee, 2020). Pri tem pa ostaja jasno, da avtomatska transkripcija ne predstavlja »končnega« besedila, temveč delovni osnutek. Raziskave post-OCR/HTR obdelave zgodovinskih in »šumnih« korpusov (angl. *noisy text*) kažejo, da tudi pri razmeroma dobrih modelih preostajajo sistematične napake, zlasti pri lastnih imenih, redkih besedah in neobičajnih črkovanjih, ki jih je treba obravnavati z dodatnimi korekcijskimi postopki (Soper, Fujimoto in Yu, 2021; Richter et al., 2018). V zadnjem času se uveljavljajo sekvenčno-sekvenčni modeli in transformerji, ki obravnavajo korekcijo kot prevajanje »šumnega« besedila v »čistejše« in pri katerih kombinacija jezikovnih modelov in jezikovno-modelskih/statističnih metod pomembno zniža delež napak (Yasin et al., 2023; Soper, Fujimoto in Yu, 2021; Chen in Ströbel, 2024). Vzporedno nastajajo pristopi za detekcijo in kvantitativno ocenjevanje rezidualnih napak, od enostavnih n-gramskih modelov za označevanje verjetno napačnih tokenov (Virk, Dannéls in Muhammad, 2021) do metod umetnega vbrizgavanja napak za oceno zanesljivosti uporabniškega pregleda (Lank, Stedman in Terry, 2010), kar omogoča, da raziskovalec hitro identificira odseke, kjer je človeški pregled nujen. Za rodoslovje iz tega sledi jasen metodološki normativ: avtomatsko transkripcijo obravnavamo kot hipotezo, nikoli kot dokaz; odločilne sklepe o identiteti in sorodstvu utemeljujemo na preverjenih odčitkih iz digitalne slike ali izvirnega dokumenta, v skladu z zahtevami GDS.

Z rodoslovnega zornega kota je bistvo OCR/HTR v dveh dopolnjujočih se funkcijah. Prvič, strojni prepis radikalno razširi »vidno polje« raziskovalca: tam, kjer je bilo prej mogoče pregledati le nekaj deset ali sto strani dnevno, je mogoče v nekaj sekundah iskati po tisočih vpisih. Nockels in sodelavci poročajo, da tolerantno iskanje po več možnih branjih (npr. Ložar/Lozar/) in delno napačno prebranih nizih bistveno poveča število relevantnih zadetkov v primerjavi s klasičnim točnim iskanjem (Nockels et al., 2022; Ströbel et al., 2020). Drugič, HTR omogoča prehod od linearne branja k strukturiranju podatkov: prepisom lahko pripnemo oznake (angl. *tags*) za osebe, kraje, datume in druge entitete, ki jih nato nadaljnja orodja (npr. tehnologija NER, ki je predstavljena v podpoglavju 4.3) uporabljajo za verjetnostno povezovanje zapisov čez različne zbirke.

V naši študiji primera se potencial OCR/HTR pokaže prav na tistih mestih, kjer smo bili še omejeni na klasično ročno delo z analognim gradivom: več ključnih virov je bilo treba poiskati in prebrati na licu mesta, brez možnosti sistematičnega pregleda celotnih serij. Če bi bilo gradivo predhodno digitalizirano in bi na reprezentativnem vzorcu nastali natančni prepisi, bi lahko v okoljih, kot sta Transkribus ali odprtokolni eScriptorium, izurili specializirane modele za posamezne serije in jih uporabili za serijsko transkripcijo celotnih zbirk, s čimer bi pridobili strojno berljiv korpus. Tak korpus bi omogočil razumno izčrpno iskanje virov (npr. vseh pojavitev priimkov Ložar/Lozar in njihovih variant ali

naslovov in parcel v Elyju, pa tudi krajevnih variant Ribnica/Ritnica/Rahnica/Reinfnica/Reifnitz), avtomatsko označevanje oseb, krajev in datumov za lažjo rekonstrukcijo družinskih konfiguracij ter preverjanje neskladij v časovnicah (npr. pri letnicah rojstva Franka mlajšega), hkrati pa bi jasno dokumentirani modeli, dosežena natančnost in dnevnik iskanj tvorili revizijsko sled, ki bi omogočala ponovljivost in transparentnost analitičnih korakov.

Pri tem ostajajo omejitve jasne. Tudi zelo dober model bo imel največ težav prav tam, kjer je rodoslovna odločitev najobčutljivejša: pri redkih priimkih, neobičajnih črkovanjih in poškodovanih vpisih. V naši študiji primera, kjer drobna razlika med »Ribnica« in »Ritnica« nosi daljnosežne posledice za sklepanje o identiteti, HTR ne more nadomestiti klasične paleografske presoje. Lahko pa jo bistveno podpre: kot filter, ki iz velikega korpusa izveleče verjetne kandidate, in kot mehanizem, ki omogoči, da standard GDS razširimo na obseg gradiva, ki je bil še pred desetletjem zaradi rokopisne oblike praktično nedosegljiv.

4.3 Prepoznavanje in normalizacija imenovanih entitet (NER)

V zadnjem desetletju se je prepoznavanje imenovanih entitet (angl. *Named Entity Recognition*, NER) v humanistiki uveljavilo kot infrastrukturna tehnologija, ki iz linearnih besedil izlušči osnovne »gradnike« zgodovinske pripovedi: osebe, kraje, organizacije, datume, denarne zneske in druge t. i. poimenovane objekte. NER-model besedilu pripiše etikete, npr. da je »Frank Lozar« oseba, »Ribnica« kraj in »15. avgust 1889« datum; iz takšnih označenih korpusov lahko nato gradimo kazala, podatkovne baze in grafe, ki povezujejo akterje, prostore in dogodke. Chastang, Torres Aguilar in Tannier (2021) to za srednjeveške latinske listine opišejo kot prehod od besedila k »matriki znanja«, v kateri je mogoče kartirati vse prostorske, socialne in osebne entitete v zbirki ter jih uporabljati kot dinamične ključne besede za nove vrste poizvedb. Podobno Ehrmann in sodelavci (2021) v preglednem članku o uporabi NER pri analizi zgodovinskih dokumentov poudarjajo, da je označevanje entitet eden prvih in ključnih korakov pri semantičnem indeksiranju obsežnih digitaliziranih zbirk. Ta korak je nenadomestljiv, če želimo iz »velikih podatkovnih zbirk preteklosti« izluščiti strukture, ki so za humanistiko relevantne. Za rodoslovje je takšen prehod skoraj intuitiven: osnovno delo rodoslovca je ravno to, da iz razpršenih virov izlušči in poveže osebe, kraje in časovne oznake v koherentne konfiguracije. NER pa to operacijo, na ravni identifikacije nespremenjenih nizov, avtomatizira in razširi, saj iz tisoče strani matičnih knjig, ladijskih manifestov, popisov prebivalstva, naturalizacijskih vlog, vojaških seznamov, časopisja in pokopaliških evidenc napravi iskalno in strukturirano podlago za nadaljnjo analizo.

Pri historičnih in večjezičnih korpusih pa se hitro pokaže, da »standardni« NER, izurjen na sodobnih angleških novicah ali Wikipediji, ne zadostuje. Ehrmann in sodelavci (2021) izpostavijo celo paleto težav: neenotno in zastarelo pravopisje, mešanje jezikov (npr. latinski, nemški in slovenski segmenti v isti zbirki), odsotnost stabilnih velikih začetnic, fonetični zapisi in napake, ki jih v besedilo vnesejo postopki OCR/HTR. Pri tem se pokaže, da zmogljivost klasičnih NER-modelov – merjena z deležem pravilno prepoznanih entitet – na historičnih besedilih pogosto močno zaostaja za rezultati modernih, »prečiščenih« korpusov. Podobne težave opisujejo tudi študije, ki so se osredotočile na druge specifične tipe zgodovinskih besedil, denimo dopisovanje, kjer se prepletajo jezikovne spremembe, ortografske variante in različne ravni neformalnosti (Won, Murrieta-Flores in Martins, 2018).

V praksi to pomeni, da NER na zgodovinskih virih ni mogoče preprosto »vklopiti«, temveč ga je treba prilagoditi. Ena možnost je učenje modelov na ročno označenih

historičnih korpusih. Chastang in sodelavci (2021) so za potrebe latinskih listin zgradili model, izurjen na korpusu srednjeveških burgundskih listin, ter pokazali, da tak specializiran pristop omogoča robustno prepoznavanje osebnih in krajevnih imen tudi v jezikovno in formalno heterogenih listinah. Álvarez de Toledo López-Herrera (2020) v poročilu o delu s španskimi umetnostnozgodovinskimi viri opiše podoben pristop: splošni model za prepoznavanje imenovanih entitet v španskem jeziku (*spaCy*) se je na zgodovinskem gradivu izkazal za nezadostnega, zato ga je na osnovi manjšega, ročno označenega korpusa ponovno izuril, razširil z dodatnimi kategorijami (datumi, denarne vsote, predmeti) in v obdelovalni tok vključil še komponento za posodobitev pravopisa. Rezultat je bila opazna izboljšava: model je po prilagoditvi zaznal bistveno več ustreznih entitet, pri osebnih in krajevnih imenih pa se je izkazal za dovolj zanesljivega, da ga je mogoče uporabljati za množično predlaganje kandidatnih zapisov – pri čemer končna presoja in potrjevanje še vedno ostaneta v rokah raziskovalca. Druga smer prilagoditve so hibridni in t. i. ensemble pristopi. Won, Murrieta-Flores in Martins (2018) pri prepoznavanju krajevnih imen v zgodovinski korespondenci pokažejo, da posamezna orodja za NER na istem gradivu pogosto dajejo različne in nestalne rezultate. Ko pa združijo več obstoječih orodij in upoštevajo tista mesta, kjer se ta ujemajo, dobijo veliko bolj stabilen in zanesljiv nabor zaznanih krajevnih imen, ki dobro zdrži tudi ob različnih načinih predobdelave in normalizacije besedila. V ozadju stoji preprosta, a pomembna ugotovitev: pri historičnih podatkih je pogosto bolj vredno, da imamo širok in razmeroma zanesljiv zajem entitet, ki raziskovalcu omogoča sistematičen pregled kandidatov, kot pa da skušamo iztisniti največ iz enega samega, na videz optimalnega modela. Na bolj splošni ravni Keraghel in sodelavci (2024) v primerjalnem pregledu NER pristopov pokažejo, da transformatorske arhitekture in VJM sicer dosegajo visoke rezultate na številnih standardnih zbirkah, a pri specializiranih ali majhnih korpusih pogosto ne presegajo, včasih pa celo zaostajajo za klasičnimi modeli (npr. CRF ali LSTM-CRF), ki so izrecno uglašeni na konkretno nalogo ali domeno. Poudarjajo, da VJM pri NER trpijo zaradi »pomanjkanja specializacije« in težjih primerov diskontinuiranih ali gnezdenih entitet ter da se kot perspektivnejša kažejo hibridna ogrodja, v katerih se široko kontekstualno znanje VJM kombinira s ciljno izurjenimi, manjšimi NER-moduli. Podobno Zhang s sodelavci (2025) na primeru pristopa *UniversalNER* pokaže, da je mogoče iz VJM s ciljno destilacijo izpeljati kompaktno modele, ki so bolj prilagojeni specifičnim zbirkam – tudi zgodovinskim in večjezičnim, npr. v kontekstu iniciative *HIPE* in historičnih jezikovnih modelov *hmbERT*.

Za rodoslovje izhajata iz teh ugotovitev predvsem dva nauka. Prvič, k osnovni nalogi NER – prepoznavanju in razvrščanju entitet – moramo vedno dodati še normalizacijo, tj. prepisovanje raznolikih zapisov v eno, dosledno referenčno obliko. V naših virih to ne pomeni le prepoznati, da sta »Frank Lozar« in »Frančišek Ložar« isti osebi, temveč tudi, da je »Ritnica«, »Rahnica«, »Reinfnica« in »Reifnitz« v danem kontekstu verjetno Ribnica na Dolenjskem, ne pa koroški Reifnitz ob Vrbskem jezeru. V praksi normalizacija prepleta več strategij: slovarčke variacij, verjetnostne modele, ki upoštevajo tipične transkripcijske napake (npr. HTR) in fonetične približke, ter kontekstualne signale (sopojavitve z drugimi znanimi entitetami). Chastang, Torres Aguilar in Tannier (2021) opozarjajo, da so v zgodovinskih jezikih ravno osebna in krajevna imena tista, kjer je variabilnost največja in kjer brez kombinacije slovarskega znanja ter strojnega učenja ni mogoče doseči zadovoljive kakovosti. Drugič, od izluščenih in normaliziranih entitet je mogoče preiti k povezovanju (angl. *entity linking*): isti Frank Lozar se pojavi v matični knjigi, na ladijskem manifestu, v popisu prebivalstva, v naturalizacijski vlogi, v vojaškem registru, v časopisni notici in na pokopališki plošči. NER tu deluje kot vhodni filter, ki iz besedil izloči vse kandidate za osebe in kraje; povezovalni postopki nato preverjajo, ali gre res za istega posameznika ali kraj, in

predlagajo verjetna ujemanja. V rodoslovju je to odločilen korak: šele ko z zadostno gotovostjo pokažemo, da je Frank Lozar z ladje SS La Bretagne ista oseba kot Frank Lozar v popisu Minnesote iz leta 1905 in kot Frank Lozar v vojaškem žrtveniku iz leta 1918, lahko verodostojno rekonstruiramo njegovo življenjsko pot in sorodstveni okvir.

V okviru GDS razumemo NER in povezovanje entitet kot orodji, ki podpirata tri temeljne faze: »razumno izčrpno iskanje« virov, analizo in korelacijo informacij ter razreševanje nasprotujočih si podatkov. Na ravni iskanja NER omogoča samodejno pripravo osebnih in krajevnih kazal čez celoten korpus, kar zmanjšuje tveganje, da bi relevantne pojavitve priimka ali toponima ostale prezrte, zlasti ob pravopisnih in jezikovnih variacijah. Na ravni analize in korelacije NER skupaj z normalizacijo zagotavlja zgoščene sezname vseh pojavitev istega ali sorodnega akterja v različnih zbirkah in zapisih; namesto ročnega prečesavanja virov lahko raziskovalec v enem koraku vidi, v katerih letih, krajih in kontekstih se pojavlja Frank Lozar, ter na tej osnovi oblikuje hipoteze o selitvenih in sorodstvenih vzorcih. Pri razreševanju nasprotij pa NER olajša prepoznavanje spornih ali protislovnih zapisov, na primer kadar se ista kombinacija osebe in približne letnice rojstva veže na različne rojstne kraje ali različne datume priselitve.

Če to preslikamo na našo študijo primera, si lahko zamislimo idealni scenarij, v katerem bi bili vsi relevantni viri, omenjeni v poglavju 2, digitizirani, zanesljivo prepisani s HTR ter obdelani s specializiranim NER in postopki normalizacije. Na tej podlagi bi NER samodejno označil vse osebe (npr. Frank/Franc/Frančišek Lozar/Ložar, Lovrenc, Neža), kraje (Prigorica, Ribnica, Tower, Ely), datume (rojstva, poroke, smrti, priselitve) in po potrebi tudi organizacije (Društvo sv. Cirila in Metoda, American Legion). Normalizacijski sloj bi poskrbel, da bi bile različne fonetične in grafične oblike priimka – Ložar, Lozar – ter kraja – Ribnica, Ritnica, Rahnica, Reinfnica, Reifnitz – prepoznane kot verjetne variante istih entitet, pri čemer bi se ohranile informacije o konkretnem zapisu in viru. Tako označen korpus bi raziskovalcu omogočil, da z eno poizvedbo izloči vse pojavitve priimka Ložar/Lozar v kombinaciji z Elyjem in okolico ter si ustvari pregled nad sorodstvenimi vezmi družine Ložar/Lozar in socialno mrežo slovensko-ameriške skupnosti v kraju. Samodejno združevanje zapisov bi olajšalo rekonstruiranje družinskih konfiguracij (npr. Frank starejši kot brat Lovrenca, strica Franka mlajšega) in omogočilo, da se v plasteh pokažejo generacijski prehodi – od pionirskega priseljenca, ki v Elyju gosti rojake in duhovnika, do vojnega heroja, po katerem je poimenovana veteranska organizacija. Hkrati bi sistematično prepoznavanje neskladij (npr. variacije rojstnih letnic in krajev pri Franku mlajšem) delovalo kot opozorilni mehanizem, ki sproži vrnitev k izvirnim vpisom in klasični kritiki virov.

Pri vsem tem je potrebna mera previdnosti. Tako kot pri OCR/HTR tudi pri NER in normalizaciji nastajajo največje napake prav pri tistih podatkih, ki so za rodoslovje najobčutljivejši: pri redkih priimkih, neobičajnih črkovanjih, poškodovanih zapisih in pri krajevnih imenih z več možnimi razlagami. Splošni NER-model lahko na primer spregleda »Frančiška« kot osebno ime ali napačno poveže »Reifnitz« s koroškim krajem namesto z Ribnico na Dolenjskem, združeni oziroma hibridni pristopi pa kljub izboljšavam ne odpravijo sistematičnih pristranskosti in napak. Keraghel in sodelavci (2024) zato opozarjajo, da je kombiniranje VJM z manjšimi, specializiranimi modeli za NER smiselno le, če so ti modeli na konkretni nalogi dobro preizkušeni in če so pragovi zaupanja ter postopki reševanja dvoumnosti jasno opisani. V rodoslovnem kontekstu se to prevede v zahtevo po jasnih pravilih normalizacije (npr. kdaj smemo »Reifnitz« razumeti kot Ribnico na Dolenjskem in kdaj ne), po izrecno določenih pragih podobnosti pri povezovanju entitet in po natančni revizijski sledi: kateri modeli so bili uporabljeni, na katerih korpusih so bili izurjeni, kakšne rezultate so dosegli na testnem naboru in katere

poizvedbe so bile izvedene v določenem času. Šele takrat lahko avtomatsko izluščene in normalizirane entitete vstopijo v rodoslovno argumentacijo kot hipoteze, ki jih preverjamo in potrjujemo z vrnitvijo k digitalni sliki ali izvirnemu dokumentu.

NER in normalizacija entitet sta v rodoslovju tako predvsem infrastrukturni orodji za širitev raziskovalčevega vidnega polja, za strukturiranje razpršenih korpusov in za odkrivanje možnih ujemanj ter nasprotij. Dokazna vrednost pa ostaja vezana na klasično kritiko virov in na GDS: šele tam, kjer se avtomatski predlogi skladajo z natančno prebranimi in v kontekstu razumljenimi zapisi v matičnih knjigah, ladijskih manifestih, popisih, vojaških in pokopaliških evidencah ipd., jih lahko sprejememo kot del verige argumentov o sorodstvenih razmerjih in selitvenih poteh. V kombinaciji s pogovornimi sistemi (poglavje 4.1) ter OCR/HTR (poglavje 4.2) NER zaokroža z UI podprt delovni tok, ki obljublja pomembno pospešitev in razširitev rodoslovnih raziskav – pod pogojem, da ostaja trdno zasidran v metodoloških normah GDS.

5 SKLEP

Prispevek, zasnovan kot metodološki premislek o rabi UI v rodoslovju, izhaja iz Genealoškega dokaznega standarda (GDS) in uveljavljene klasične metode dela, na kateri nato preizkušamo možnosti z UI podprtih pristopov. Na primeru družine Ložar/Lozar smo najprej strnjeno rekonstruirali klasični rodoslovni postopek: od razumno izčrpnega iskanja virov in natančnega navajanja ter kritične analize do razreševanja nasprotij in oblikovanja sklepa o sorodstvenih razmerjih. Ta rekonstrukcija ni bila cilj sama po sebi, temveč je služila kot referenčni okvir, na katerega smo v nadaljevanju navezali razpravo o vlogi orodij UI: šele ko jasno vidimo, kako GDS usmerja »ročni« potek raziskave, lahko smiselno razmišljamo o tem, kje in pod kakšnimi pogoji ga je mogoče dopolniti z novimi tehnologijami.

Na tej osnovi smo pogovorne sisteme, tehnologije OCR/HTR (optično prepoznavanje znakov in strojno prepoznavanje rokopisov) ter NER (prepoznavanje imenovanih entitet) obravnavali kot tri plasti iste infrastrukturne podpore. Pogovorni sistemi lahko, skladno z obravnavano literaturo in študijo primera družine Ložar/Lozar, pomagajo pri orientaciji po razpršenih korpusih virov, generiranju variant imen in toponimov ter pri oblikovanju raziskovalnih strategij, ne da bi postali vir sami po sebi. Tehnologiji OCR/HTR omogočata prehod od rokopisnih in tiskanih podob k strojno berljivemu besedilu ter s tem bistveno širita obseg gradiva, znotraj katerega je sploh mogoče govoriti o »razumno izčrpnem iskanju«. NER z normalizacijo in povezovanjem entitet iz teh besedil gradi kazala in mreže, ki olajšajo sistematično iskanje, korelacijo informacij in prepoznavanje neskladij v časovnicah.

V obravnavani literaturi in na podlagi naše študije primera se kot skupen metodološki poudarek kaže predvsem dvoje. Prvič, orodja UI lahko rodoslovno delo občutno pospešijo in razširijo: povečajo dostopnost virov, izboljšajo sledljivost raziskovalnih korakov in odpirajo nove analitične poglede na selitvene in sorodstvene mreže. Drugič, skoraj v vseh študijah se ponavlja opozorilo, da morajo takšni postopki ostati vpeti v jasno opredeljene standarde dokazovanja, transparentno dokumentirane postopke normalizacije in premišljene prakse ravnanja z občutljivimi osebnimi in družinskimi podatki. Naša študija primera tem opozorilom pritrjuje: prav pri redkih priimkih, variabilnih črkovanjih in večjezičnih fonetičnih zapisih so napake orodij UI hkrati najverjetnejše in za rodoslovne sklepe najnevarnejše.

V tem smislu zagovarjamo hibridni model rodoslovnega dela: klasična metoda in GDS ostajata okvir, v katerega so vpeti pogovorni sistemi, OCR/HTR in NER. S

prispevkom ne ponujamo tehnološke zamenjave uveljavljenih (klasičnih) metod v rodoslovju, temveč raziskujemo načine, kako orodja UI umeščati kot infrastrukturno dopolnitev, ne kot nadomestilo rodoslovne presoje. Prihodnje raziskave bodo morale ta okvir empirično preizkusiti na širšem naboru primerov, razviti domensko prilagojene modele za slovensko (izseljensko) gradivo ter oblikovati skupnostne smernice, ki bodo rabo UI v rodoslovju utrdile kot metodološko in etično vzdržno prakso.

6 VIRI IN LITERATURA

VIRI

EWHMA – Ely-Winton History Museum Archives (Ely (Minnesota)):

Spominska diploma na Franka Lozarja. World Wars 1 + 2, John J. Pershing. Fond Large Military photos – documents, EWHM.

NALJ – Nadškofijski arhiv Ljubljana (Ljubljana, Slovenija):

Krstna knjiga za obdobje 1886–1896, str. 140: Frančišek Ložar (22. 10. 1894) – Ribnica 100, Fond Župnija Ribnica, Nadškofijski arhiv Ljubljana

NARA – National Archives and Records Administration (Washington, D. C. in College Park (Maryland)):

Manifest ladje S.S. La Bretagne, 26. 5. 1900–3. 6. 1900. Serija mikrofilmov T715 in M237, zvezek 129, fotografija 253 od 704, NARA.

Passenger and crew lists of vessels arriving at New York, 1897–1942; index to passenger lists of vessels arriving at New York, June 16, 1897– June 30, 1902, Agnes Lozar. Serija mikrofilmov T519, zvezek 70, NARA.

ČASNIKI

Ameriška domovina – Cleveland (Ohio), 1933.

Peshel, P. P. (1933a). Na Ely, Minn., je preminul najstarejši Slovenec. *Ameriška domovina*, 35 (1933) (12), 16. 1. 1933, 2.

Peshel, P. P. (1933). Zanimive vesti iz življenja naših ljudi po ameriških naselbinah. *Ameriška domovina*, 35 (1933) (9), 12. 1. 1933, 1.

Glas naroda – New York City (New York), 1921.

Pogorelec, M. (1921). S pota. *Glas naroda*, 29 (1921), 163, 14. 7. 1921, 4.

Glasilo K.S.K. Jednote – (Joliet), 1921.

Poročevalec. (1921). Društvena naznanila in dopisi. Ely, Minn. *Glasilo K.S.K. Jednote*, 28 (1921) (28), 13. 7. 1921, 2.

Proletarec – Chicago (Illinois), 1933.

Uredništvo. (1933a). Frank Lozar, najstarejši slovenski priseljenec, umrl. *Proletarec*, 28 (1933) (1323), 19. 1. 1933, 3.

Prosveta – Chicago (Illinois), 1933.

Uredništvo. (1933). Domač drobiž. Še o smrti najstarejšega ameriškega Slovenca. *Prosveta*, 26 (1933) (6), 10. 1. 1933, 2.

The Ely Miner – Ely (Minnesota), 1919, 1921.

Uredništvo. (1919). Legion Meeting is Interesting. Legion Perfects Organization Monday Evening and Transacts much Business. *The Ely Miner*, 25 (1919) (18), 7. 11. 1919, 1.

Uredništvo. (1921). Body to Arrive. Mother Receives Message that Body of War Hero Frank Lozar Had Arrived at New York. *The Ely Miner*, 26 (1921) (50), 17. 6. 1921, 1.

SPLETNI VIR

Lozar, Frank, *Family Search*. Digitalni arhiv Family Search, United States World War I Draft Registration Cards, 1917–1918, Minnesota, St. Louis County, no. 3, A – M, Registration Card No. 199. dostopno na: <https://www.familysearch.org/ark:/61903/3:1:33S7-818P-QYS?i=1389&wc=9FZZ-3TP%3A928311201%2C928767601&cc=1968530>, pridobljeno: 29. 7. 2025.

LITERATURA

- Altmäe, S., Sola-Leyva, A., Salumets, A. (2023). Artificial Intelligence in Scientific Writing: A Friend or a Foe? *Reproductive BioMedicine Online*, 47, 3–9. <https://doi.org/10.1016/j.rbmo.2023.04.009>.
- Álvarez de Toledo López-Herrera, F. (2020). Automated Tagging of Historical, Non-English Sources with Named Entity Recognition (NER): A Resource. *Center for Data and Visualization Sciences, Duke University Libraries*, dostopno na: <https://blogs.library.duke.edu/data/2020/08/31/automated-tagging-of-historical-non-english-sources-with-named-entity-recognition-ner-a-resource/>, pridobljeno: 17. 8. 2025.
- Bluche, T., Messina, R. (2017). Gated Convolutional Recurrent Neural Networks for Multilingual Handwriting Recognition. *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 646–651, <https://doi.org/10.1109/ICDAR.2017.111>.
- Board for Certification of Genealogists. (2021). *Genealogy Standards*, 2nd edition. Nashville: Ancestry.
- Buholayka, M., Zouabi, R., Tadinada, A. (2023). Is ChatGPT Ready to Write Scientific Case Reports Independently? A Comparative Evaluation Between Human and Artificial Intelligence. *Cureus*, 15, e39386. <https://doi.org/10.7759/cureus.39386>.
- Carretero, M., Gartner, E. (2024). Artificial Intelligence and historical thinking: a dialogic exploration of ChatGPT / Inteligencia Artificial y pensamiento histórico: una exploración dialógica del ChatGPT. *Studies in Psychology*, 45 (1), 80–102. <https://doi.org/10.1177/0210939524124137>.
- Chastang, P., Torres Aguilar, S. O., Tannier, X. (2021). A Named Entity Recognition Model for Medieval Latin Charters. *Digital Humanities Quarterly*, 15 (4). <https://hdl.handle.net/10993/59645>.
- Chiarello, F., Giordano, V., Spada, I., Barandoni, S. & Fanton, G. (2024). Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation*, 133, 103002. <https://doi.org/10.1016/j.technovation.2024.103002>.
- Currie, G. M. (2023). Academic integrity and artificial intelligence: Is ChatGPT hype, hero or heresy? *Seminars in Nuclear Medicine*, 53, 719–730. <https://doi.org/10.1053/j.semnuclmed.2023.04.008>.
- Deng, J. & Lin, Y. (2023). The benefits and challenges of ChatGPT: an overview. *Frontiers in Computing and Intelligent Systems*, 2, 81–83. <https://doi.org/10.54097/fcis.v2i2.4465>.
- Drinkall, F., Rahimikia, E., Pierrehumbert, J., & Zohren, S. (2024). Time Machine GPT. In Duh, K., Gomez, H., & Bethard, S. (Eds.), *Findings of the Association for Computational Linguistics: NAACL 2024*, 3281–3292. <https://doi.org/10.18653/v1/2024.findings-naacl.208>.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M. et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>.
- Ehrmann, M., Hamdi, A., Pontes, E. L., Romanello, M., & Doucet, A. (2021). Named Entity Recognition and Classification in Historical Documents: A Survey. *ACM Computing Surveys* 56 (2), 1–47. <https://doi.org/10.1145/360493>.
- Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, 613, 423. <https://doi.org/10.1038/d41586023-00056-7>.
- Friš, D. (1994). Jugoslovanska katoliška jednota (1898–1920). *Dve domovini/Two Homelands: razprave o izseljenstvu*, 5, 37–62.
- Gîrbacia, F. (2024). An Analysis of Research Trends for Using Artificial Intelligence in Cultural Heritage. *Electronics*, 13 (18), 3738. <https://doi.org/10.3390/electronics13183738>.

- Gonzalez Garcia, G & Weilbach, W. (2023).** If the Sources Could Talk: Evaluating Large Language Models for Research Assistance in History. *arXiv*. Dostopno na: [arXiv:2310.10808](https://arxiv.org/abs/2310.10808) (pridobljeno: 2. April 2025).
- Green, S. (2024).** The Genealogical Proof Standard (GPS): A Review. *Legacy Family Tree Webinars*, dostopno na: https://familytreewebinars.com/wp-content/uploads/2023/12/Syllabus_ShannonGreen_GPS_16Jan2024.pdf, pridobljeno: 22. 6. 2025.
- Hazemali, D. (2023).** Safeguarding liberty? Repressive measures against enemy aliens and community resilience in WWI United States : the Slovenian-American experience. *Annales : anali za istrske in mediteranske študije. Series historia et sociologia*, 33 (3), 489–502. <https://doi.org/10.19233/ASHS.2023.24>.
- Hazemali, D., Osojnik, J., Onič, T., Todorović, T. & Borovič, M. (2024).** Evaluating Chatbot Assistance in Historical Document Analysis. *Moderna arhivistika*, 7 (2), 53–83. <https://doi.org/10.54356/ma/2024/biub3010>.
- Hendriks, B., Groth, P., & van Erp, M. (2021).** Recognising and linking entities in old dutch text: A case study on voc notary records. *CEUR Workshop Proceedings* 2810, 25–36. <https://ceur-ws.org/Vol-2810/paper3.pdf>.
- Ingle, R. R., Fujii, Y., Deselaers, T., Baccash, J. & Popat, A. C. (2019).** A Scalable Handwritten Text Recognition System. *2019 International Conference on Document Analysis and Recognition (ICDAR)*, Sydney, NSW, Australia. 17–24. <https://doi.org/10.1109/ICDAR.2019.00013>.
- Jaillant, L. & Aske, K. (2023).** Are Users of Digital Archives Ready for the AI Era? Obstacles to the Application of Computational Research Methods and New Opportunities. *Journal on Computing and Cultural Heritage* 16 (4), 87. <https://doi.org/10.1145/3631125>.
- Jaillant, L. (Ed.) (2022).** *Archives, access and artificial intelligence: Working with born-digital and digitized archival collections*. Bielefeld: Bielefeld University Press. <https://hdl.handle.net/2134/19687587.v1>.
- Jones, T. W. (2013).** *Mastering Genealogical Proof*. Arlington: National Genealogical Society.
- Kansteiner, W. (2022).** Digital Doping for Historians: Can History, Memory, and Historical Theory be Rendered Artificially Intelligent? *History and Theory*, 61, 119–133. <https://doi.org/10.1111/hith.12282>.
- Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., Gershovits, M., Markus, B., Sheikh, M., Gymrek, M., Bhatia, G., MacArthur, D. G., Price, A. L., & Erlich, Y. (2018).** Quantitative analysis of population-scale family trees with millions of relatives. *Science*, 360, 6385, 171–175. <https://doi.org/10.1126/science.aam9309>.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E. et al. (2023).** ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>.
- Keraghel, I., Morbieu, S., & Nadif, M. (2024).** A survey on recent advances in named entity recognition. *arXiv*. Org, abs/2401.10825. <https://doi.org/10.48550/arxiv.2401.10825>.
- Khan, N., Ab Rahman, A. A., Rizvi, S. M., Alshareef, I. Y., Marsono, M. N., & Bakht, M. P. (2025).** Systematic Literature Review of Machine Learning Models and Applications for Text Recognition. *IEEE Access*, 13, 177647–177670. <https://doi.org/10.1109/ACCESS.2025.3618109>.
- Khudair, G. H. (2025).** Artificial Intelligence Historiography: Extensive Explorations in Reshaping Modern History and Social Studies. In García Márquez, F. P. et al. (Eds), *Computing, Internet of Things and Data Analytics. ICCIDA 2024. Smart Innovation, Systems and Technologies*, vol 387. https://doi.org/10.1007/978-3-031-87154-2_39.
- Liang, S., Zhu, B., Zhang, Y., (...), Jin, J. (2020).** *Proceedings - 2020 IEEE 22nd International Conference on High Performance Computing and Communications*.
- Liu, Z., Song, R., Li, K., & Li, Y. (2025).** From Detection to Understanding: A Systematic Survey of Deep Learning for Scene Text Processing. *Applied Sciences*, 15 (17), 9247. <https://doi.org/10.3390/app15179247>.
- Lozić, E. & Štular, B. (2023).** Fluent but Not Factual: A Comparative Analysis of ChatGPT and Other AI Chatbots' Proficiency and Originality in Scientific Writing for Humanities. *Future Internet*, 15, 336. <https://doi.org/10.3390/fi15100336>.

- Matjaž Klemenčič, M. & Šeruga, T. (2019).** *Pregled zgodovine slovenske skupnosti v Elyju, Minnesota*. Maribor, Bielsko-Biala, Budapest, Kansas, & Praha: Univerzitetna založba Univerze v Mariboru.
- Mills, E. S. (1997).** *Evidence! Citation & Analysis for the Family Historian*. Baltimore: Genealogical Publishing Company.
- Muehlberger, G., Seaward L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinöcker, A., Grüning, T., Hackl, G., Haukkovaara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, M., Kahle, P., Kallio, M., Kaplan, F., Kleber, F., Labahn, R., Lang, E. M., Laube, S., Leifert, G., Louloudis, G., McNicholl, R., Meunier, J., Michael, J., Mühlbauer, E., Philipp, N., Pratikakis, I., Puigcerver Pérez, J., Putz, H., Retsinas, G., Romero, V., Sablatnig, R., Sánchez, J. A., Schofield, P., Sfikas, G., Sieber, C., Stamatopoulos, N., Strauß, T., Terbul, T., Toselli, A. H., Ulreich, B., Villegas, M., Vidal, E., Walcher, J., Weidemann, M., Wurster, H., & Zagoris, K. (2019).** Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation* 75 (5), 954–976. <https://doi.org/10.1108/JD-07-2018-0114>.
- Neto A. F. D. S., Bezerra B. L. D., Araújo S. S., Souza W. M. A. S., Alves K. F., Oliveira M. F., Lins S. V. S., Hazin H. J. F., Rocha P. H. V., & Toselli A. (2024).** H. BRESSAY: A Brazilian Portuguese Dataset for Offline Handwritten Text Recognition. *Lecture Notes in Computer Science*. 315–333. https://doi.org/10.1007/978-3-031-70536-6_19.
- Nockels J., Gooding, P., Ames, S., & Terras, M. (2022).** Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research. *Archival science*, 22 (3), 367–392. <https://doi.org/10.1007/s10502-022-09397-0>.
- Palladino, F. (2025).** Intelligenza artificiale generativa e fonti storico-educative: prospettive metodologiche. *History of Education and Children's Literature*, 20 (1), 537–550. <https://doi.org/10.48219/1387>.
- Ramjee, D. & Ringrose, K. N. (2020).** The Challenges of Forensic Genealogy: Dirty Data, Electronic Evidence, and Privacy Concerns. *Denver Law Review*, 98 (1). <https://ssrn.com/abstract=3780373>.
- Ray, P. P. (2023).** ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>.
- Stelter, R., & Biehler, R. (2025).** Data retrieval from local heritage books—Is artificial intelligence the solution? *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 1–19. <https://doi.org/10.1080/01615440.2025.2512744>.
- Stokel-Walker, C. (2023).** ChatGPT listed as author on research papers: many scientists disapprove. *Nature*, 613, 620–621. <https://doi.org/10.1038/d41586-023-00107-z>.
- Ströbel, P. B., Clematide, S., & Volk, M. (2020).** How Much Data Do You Need? About the Creation of a Ground Truth for Black Letter and the Effectiveness of Neural OCR. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 3551–3559. <https://www.aclweb.org/anthology/2020.lrec-1.436>.
- Ströbel, P.B., Hodel, T., Boente, W., Volk, M. (2023).** The Adaptability of a Transformer-Based OCR Model for Historical Documents. In Coustaty, M., Fornés, A. (Eds.), *Document Analysis and Recognition – ICDAR 2023 Workshops*. ICDAR 2023. Lecture Notes in Computer Science, vol. 14193. https://doi.org/10.1007/978-3-031-41498-5_3.
- Tirado-Olivares, S., Navío-Inglés, M., O'Connor-Jiménez, P. & Cózar-Gutiérrez, R. (2023).** From Human to Machine: Investigating the Effectiveness of the Conversational AI ChatGPT in Historical Thinking. *Education Sciences*, 13, 803. <https://doi.org/10.3390/educsci13080803>.
- Trunk, J. M. (1912).** *Amerika in Amerikanci*. Celovec: samozaložba.
- Vargas-Murillo, A. R., Sotelo-Calderon, A. F., Gómez-Zegarra, J. L., & Zegarra-Ponce, L. R. (2024).** The Role of Artificial Intelligence and Pattern Recognition in the Authentication and Analysis of Historical Documents: A Literature Review. In Ranganathan, G., Papakostas, G. A. & Shi Y. (Eds.), *Inventive Communication and Computational Technologies - Proceedings of ICICCT 2024*. Lecture Notes in Networks and Systems; Vol. 23 LNNS, 759–768). https://doi.org/10.1007/978-981-97-7710-5_58.
- Won, M., Murrieta-Flores, P., & Martins, B. (2018).** Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora. *Frontiers in Digital Humanities* 5 (2). <https://doi.org/10.3389/fdigh.2018.00002>.

- Wu, X., Jiang, T., Zhu, Y., & Bu, C. (2023). Knowledge Graph for China's Genealogy. In *IEEE Transactions on Knowledge and Data Engineering*, 35 (1), 634–646, <https://doi.org/10.1109/TKDE.2021.3073745>.
- Zaplotnik, J. L. (1969). Monsinjur Franc Jožef Buh: Misijonar v Minnesoti. *Ave Maria koledar* 57 (10), 49–84.
- Zhang, S., & Colavizza, G. (2025). Named Entity Recognition of Historical Text via Large Language Model. *arXiv.Org*, arXiv:2508.18090. <https://doi.org/10.48550/arXiv.2508.18090>.

SUMMARY

THE POTENTIAL OF ARTIFICIAL INTELLIGENCE IN GENEALOGY: A METHODOLOGICAL REFLECTION USING THE CASE STUDY OF THE SLOVENIAN EMIGRANT FAMILY LOŽAR/LOZAR FROM ELY, MINNESOTA

doc. dr. David HAZEMALI

Assistant professor and research fellow, University of Maribor, Faculty of Arts, Department of History, Slovenia
david.hazemali@um.si

mag. Nina GOSTENČNIK

Archival adviser, directress, Regional Archives Maribor, Slovenia
nina.gostencnik@pokarh-mb.si

asist. dr. Mladen BOROVIČ

Teaching assistant and researcher, Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Slovenija
mladen.borovic@um.si

The article, conceived as a methodological reflection on the use of AI in genealogy, takes the Genealogical Proof Standard (GPS) and established classical practice as its point of departure. Using the Ložar/Lozar family as a case study, we first reconstruct, in condensed form, the classical genealogical workflow: from reasonably exhaustive research and precise citation through critical source analysis to the resolution of conflicts and the formulation of a conclusion about kinship relations. This reconstruction was not an end in itself but served as a reference framework for the subsequent discussion of AI-supported tools: only once the GPS-guided “manual” research process is clearly outlined can we meaningfully consider where, and under what conditions, it can be complemented by new technologies.

On this basis, we treat chatbots, OCR/HTR technologies (optical character recognition and handwritten text recognition), and NER (named entity recognition) as three layers of the same infrastructural support, and we illustrate their potential on the Ložar/Lozar case study. In this context, chatbots can, in line with the literature discussed, assist in navigating dispersed corpora of sources, generating variants of names and toponyms, and shaping research strategies, without becoming a source in their own right.

OCR/HTR enable the transition from handwritten and printed pages to machine-readable text and thus significantly expand the volume of material within which “reasonably exhaustive” research is even possible. NER, combined with normalisation and entity linking, builds indices and networks from these texts that facilitate systematic searching, correlation of information, and the detection of inconsistencies in timelines.

Across the literature considered, and on the basis of our case study, two methodological emphases emerge. First, AI-supported tools can substantially accelerate and extend genealogical work: they increase access to sources, improve the traceability of research steps, and open up new analytical perspectives on migratory and kinship networks. Second, nearly all studies reiterate that such procedures must remain embedded in clearly defined standards of proof, transparently documented normalisation workflows, and carefully considered practices for handling sensitive personal and family data. Our case study confirms these cautions: precisely for rare surnames, variable spellings, and multilingual phonetic renderings, the errors of AI tools are both most likely and most consequential for genealogical conclusions.

In this sense, we advocate a hybrid model of genealogical research: classical methods and the GPS remain the framework into which chatbots, OCR/HTR, and NER are inserted. Rather than offering a technological replacement for established (classical) genealogical methods, this article explores ways of positioning AI tools as infrastructural supplements rather than substitutes for genealogical judgement. Future research will need to test this framework empirically across a broader range of cases, develop domain-adapted models for Slovenian (emigrant) materials, and formulate community guidelines that can consolidate the use of AI in genealogy as a methodologically and ethically robust practice.

O avtorjih:

Dr. David Hazemali je docent za sodobno zgodovino in znanstveni sodelavec na Oddelku za zgodovino Filozofske fakultete Univerze v Mariboru. Njegovo akademsko področje vključuje selitvene in etnične študije, vojaško zgodovino 20. stoletja, jugoslovansko in sodobno slovensko zgodovino, jezik diplomacije ter inovativno rabo pogovornih sistemov umetne inteligence v humanistiki. Trenutno se osredotoča predvsem na podoktorski projekt o izkušnjah južnih Slovanov iz nekdanje Avstro-Ogrske, ki so med prvo svetovno vojno živeli v Združenih državah Amerike. Dejavno sodeluje v interdisciplinarnih raziskovalnih projektih, ki raziskujejo kulturne, družbene in okoljske spremembe, svoje ugotovitve pa redno objavlja v vodilnih recenziranih revijah in predstavlja na mednarodnih ter domačih konferencah, okroglih mizah in simpozijih.

Mag. Nina Gostenčnik, rojena v Mariboru, je končala podiplomski študij na Filozofski fakulteti Univerze v Mariboru. Leta 2006 se je zaposlila v Pokrajinskem arhivu Maribor kot arhivistka za področje gospodarstva in do leta 2017 pridobila strokovni naziv arhivska svetovalka. Od leta 2021 je direktorica Pokrajinskega arhiva Maribor. Je urednica revije Moderna arhivistika in serij Gradivo za zgodovino Maribora ter Gradivo za zgodovino Koroške in članica več uredniških odborov in strokovnih komisij. Kot predavateljica je sodelovala na številnih mednarodnih in nacionalnih znanstvenih posvetih ali strokovnih srečanjih. Je avtorica več kot 170 objavljenih ali izvedenih del s področja arhivske teorije in prakse, zgodovinopisja in implementacije družbenih omrežij na področju arhivske dejavnosti.

Dr. Mladen Borovič je asistent in raziskovalec na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru. Njegova raziskovalna področja so aplikativna umetna inteligenca, iskalniki in priporočilni sistemi, obdelava naravnega

jezika, detekcija podobnih vsebin, visokozmogljivo računalništvo in generativna umetna inteligenca. Aktivno deluje na projektih, povezanih z obdelavo naravnega jezika in uporabo umetne inteligence v obliki pogovornih sistemov. Sodeloval je pri odmevnejših projektih, kot so Nacionalna infrastruktura odprtega dostopa, Slovenščina na dlani, HPC RIVR in Razvoj slovenščine v digitalnem okolju. V sklopu svojega pedagoškega dela poučuje predmete, povezane z računalniškimi omrežji, visokozmogljivim računalništvom in umetno inteligenco. Je tudi recenzent za več znanstvenih revij s faktorjem vpliva na področju računalništva.

About the authors:

David Hazemali, Ph. D., is an assistant professor and research fellow at the Department of History, Faculty of Arts, University of Maribor, Slovenia. His academic expertise includes migration and ethnic studies, 20th-century military history, Yugoslav and contemporary Slovenian history, the language of diplomacy, and the innovative use of conversational AI in the humanities. His current primary focus is a postdoctoral project on the experiences of South Slavs from the former Austro-Hungarian Empire living in the United States during the First World War. He is actively involved in interdisciplinary research projects on cultural, social, and environmental transitions, and has published in leading peer-reviewed journals as well as presented his work at international and local conferences and symposia.

Nina Gostenčnik, M. Sc., completed her postgraduate studies at the Faculty of Arts, University of Maribor. In 2006, she began working at the Regional Archives Maribor as an archivist specializing in the field of economics. By 2017, she had earned the professional title of archival advisor. Since 2021, she has been the director of the Regional Archives Maribor. She is the editor of the journal *Moderna arhivistika* and the series *Materials for the History of Maribor* and *Materials for the History of Carinthia*. Additionally, she is a member of several editorial boards and professional committees. As a lecturer, she has participated in numerous international and national scientific conferences and professional meetings. Nina Gostenčnik is the author of more than 170 published or presented works in the fields of historiography, archival theory and practice, and the implementation of social media in archival activities.

Mladen Borovič, Ph. D., is a teaching assistant and researcher at the Faculty of Electrical Engineering and Computer Science at the University of Maribor. His research areas include applied artificial intelligence, search engines and recommender systems, natural language processing, similar content detection, high-performance computing, and generative artificial intelligence. He actively works on projects related to natural language processing and the application of artificial intelligence in the form of conversational systems. He has participated in notable projects such as the National Open Access Infrastructure, *Slovene in the Palm of Your Hand*, HPC RIVR, and the *Development of Slovene in a Digital Environment*. As part of his teaching, he lectures on topics related to computer networks, high-performance computing, and artificial intelligence. He is also a reviewer for several peer-reviewed scientific journals with an impact factor in the field of computer science.