

ELEKTRONSKE ZNANSTVENOKRITIČNE IZDAJE SLOVENSKEGA SLOVSTVA eZISS: METODE ZAPISA IN IZDAJE

Tomaž Erjavec

Institut "Jožef Stefan", Ljubljana

Matija Ogrin

Inštitut za slovensko literaturo in literarne vede, ZRC SAZU, Ljubljana

UDK 821.163.6:801.73:004.91

Knjižnica znanstvenih izdaj eZISS povezuje klasično metodologijo filološko-kritičnega izdajanja besedil z modernimi tehnološkimi postopki, ki danes nudijo raznoliko podporo znanstveni obravnavi humanističnih besedil. Z naslonitvijo na mednarodne standarde elektronskega označevanja besedil je bilo moč pripraviti v okviru eZISS več izdaj temeljnih del slovenske literature. Prispevek oriše nekatere postopke, ki so to omogočili.

znanstvenokritične izdaje, slovenska književnost, ekdotika, tekstna kritika, XML, TEI

The eZISS library of scholarly editions links the traditional methodology of philological critical editing with modern technology, which offers varied kinds of support to the scholarly investigation of texts. The application of international standards for text encoding allowed us to publish electronic text-critical editions of a number of historically important works of Slovene literature. The paper outlines some of the approaches that made this possible.

scholarly editing, critical editions, Slovene literature, textual criticism, XML, TEI

1 Metodološka izhodišča

1.1 Namen in cilj

Temeljni namen projekta eZISS (<http://nl.ijs.si/e-zrc/>), katerega prvi nastavki segajo v leto 2001 (Ogrin, 2005), sta bila oživitve zanimanja za teorijo in prakso znanstvenega izdajanja literarnih besedil v znanstvenokritičnih izdajah ter prenos te tradicionalne teorije in prakse v novo okolje – v elektronski medij. Zdelo se je, da je bilo glede na pomembnost znanstvenih izdaj v sistemu literarne vede na Slovenskem obnovljeno zanimanje zanje potrebno, še posebej za del, ki zadeva prenos tradicionalnih ekdotičnih (ali tekstnokritičnih) metod iz medija tiskane knjige v elektronsko izdajo. Izkoristiti smo želeli možnosti novih informacijskih tehnologij, med katerimi so nekatere specifično prilagojene prav besedilom, še več, humanističnim besedilom, in z njihovo pomočjo pripraviti sodobne elektronske znanstvenokritične izdaje. Takšna izdaja

naj se odlikuje predvsem po tem, da za analizo in predstavitev konkretnega besedila uporabi vse tiste ekdotične metode, ki so za sporočilo tega besedila relevantne ali primerne. Ali povedano z uporabniške strani: izdaja naj predstavi slovstveni vir v vseh prikazih, ki so lahko zanimivi raznolikim skupinam bralcev in znanstvenikov. To konkretno pomeni, da bo izdaja bralcu, ki ga nekoliko zanima tudi zgodovina slovenskega jezika, primerno odgovorila z diplomatičnim prepisom, torej takim, ki ohranja čim več prvih izvirnika; nekomu, ki ga bolj zanimajo narečne prvine, bo največ dal fonetični prepis istega vira; kritični prepis v sodobni pravopisni normi bo najbolj ustregel »navadnemu bralcu« (toda, kdo natanko je to?); arhivarja in zgodovinarja pa bo najbolj pritegnil digitalni faksimile tega vira – pa tudi vsak »navadni« bralec si ga z zanimanjem ogleda, saj instinktivno čuti, da se v faksimilu skriva tisti pravi, primarni vir besedila, iz

katerega prejema vse drugo v izdaji svojo avtentičnost in pomen.

Skratka: v literarni vedi na Slovenskem se zaradi znanih razlogov skorajda ne more zgoditi, da bi neki slovenist pripravil določeno izdajo besedila po neki (eni) metodi, drugi slovenist pa bi to isto besedilo izdal čez nekaj let – kot dopolnilo – še enkrat, toda po drugi metodi (ali več metodah). Nasprotno. Kar znanstveno izdamo, gotovo dolga desetletja ne bo ponovno izdano. To je praktični (celo pragmatični) argument za to, da bi vir besedila, ko ga že pripravljamo za znanstveno objavo, proučili, prepisali in pripravili za objavo *po več kot le eni metodi*. Metodološki in epistemološki argument za to pa je, da se za isto besedilo (morda še zlasti dandanes) razne skupine bralcev zanimajo na različne načine in bi v svoj izkustveni interes želele zajeti isti vir, isto besedilo v različnih perspektivah – če jim le izdaja to omogoča. Nedvomno gre v vseh primerih za isto besedilo, toda fenomen njegovega prikaza lahko omogoča razne prvine razumevanja, od jezikovno-historičnih in narečnih do vsebinskih.

1.2 Gradivo in stanje izdajanja

Na teh metodoloških izhodiščih je potekalo prizadevanje za pripravo izdaj eZISS. Doslej je izšlo sedem takšnih izdaj, to je povprečno po ena na leto:

1. *Brižinski spomeniki*. (Izdaja, osnovana na Bernik idr., 1993, je svojevrsten *editio variorum*, obsega najbogatejši metodološki razpon: faksimile, več starejših in sodoben (Igor Grdina) diplomatični prepis, dva kritična prepisa, fonetični prepis (Tine Logar), prevode v več jezikov, integriran zvočni posnetek, slovar besedja, študije itn.)

2. *Škofjeloški pasijon*. (Faksimile, diplomatični, kritični prepis, dopolnilno dokumentarno gradivo, integriran video posnetek uprizoritve, študije.)

3. *Mestna prisežna besedila*. (36 slovenskih priseg, nastalih od 17. do začetka 19.

stol., predstavljenih s faksimilom, diplomatičnim in kritičnim prepisom; slovarček, študije.)

4. *Korespondenca Žige Zoisa*. (Faksimile, diplomatični prepis, prevod v slovenščino, študije.)

5. Anton Martin Slomšek: *Tri pridige o jeziku*. (Faksimile, diplomatični, kritični prepis, študije.)

6. Izidor Cankar: *S poti*. (Dve verziji slovenskega esejističnega romana, diplomatični prepis, kritični aparat, ki evidentira čez 400 razlik med obema verzijama, študije.)

7. Anton Podbevšek: *Zbrane pesmi*. (Prva izdaja, ki prinaša celotno delo slovenskega avantgardista, faksimile, diplomatični, kritični prepis, študije.)

Gradivo ali primarni viri, iz katerih potekajo izdaje, so zelo raznoliki: segajo od srednjeveškega (BS) in baročnega (ŠP) rokopisa do pisemske korespondence, prvotiskov v 19. stoletju in že izdanih, toda tako ali drugače avtorsko revidiranih modernih besedil 20. stoletja. Vsak od teh virov je narekoval svojstven način obravnave, upoštevajoč njegovo historično, jezikovno, zvrstno in avtorsko specifičnost. V tem prispevku se ni mogoče posvetiti posameznim ekdotičnim metodam in edicijskim postopkom, uporabljenih v teh izdajah, vsekakor pa je bil pri tem cilj vsake izdaje verodostojno in kritično vsestransko predstaviti vir kot tak. V nadaljevanju zato nekoliko orišemo uporabljeno tehnološko podlago, ki je bila dovolj univerzalna, da je mogla adekvatno izpolniti vse, tudi najtežje zahteve izdaj eZISS.

2 Tehnološka platforma in dostopnost

Postopek priprave digitalnih izdaj je bil sicer odvisen od specifik vsake posamezne izdaje, vendar je bilo vsem skupno, da je bilo izvorno besedilo najprej pripravljeno v digitalni obliki, iz te polavtomatsko pretvorjeno v standardni zapis v obliki XML (Extensible Markup Language), ta pa avtomatsko v zapis

HTML, primeren za spletno branje. V prispevku ne bomo opisovali korakov pretvorb izvirnih besedil v XML, saj se je pristop, v odvisnosti od posebnosti izvirnega formata in želenega zapisa v XML, razlikoval od izdaje do izdaje, pač pa v nadaljevanju tega razdelka podrobneje raje opišemo tehnološka načela, ki so nas vodila pri pripravi edicij eZISS.

2.1 Kanonični zapis izdaj v XML

Brez standardiziranega zapisa digitalnih besedil se njihova trajnost ter izmenljivost bistveno zmanjšata. Razvoj elektronskega objavjanja besedil je zato naravnano k izdelavi in uporabi standardiziranega kodiranja in označevanja besedil (mark-up), ki naj nevtralizira vpliv spreminjajoče se strojne in programske opreme na elektronske tekste. Takšen standard je označevanje besedil v jeziku XML, ki ga predlaga Text Encoding Initiative (TEI) s svojimi obsežnimi in podrobnimi priporočili, ki so v celoti dostopna na medmrežju, na spletni strani konzorcija TEI, <http://www.tei-c.org/>.

Priporočila TEI sestavljajo smernice, tj. prozni opisi oznak, in formalni del, tj. vrste posameznih modulov, ki jih je mogoče kombinirati, da bi ustvarili željeno shemo XML, ustrezno namenom določenega projekta. S takšno shemo je nato, npr. v urejevalniku za XML, možno formalno validirati zapis XML posameznih e-izdaj, obenem pa je dostopna tudi dokumentacija za vse uporabljene elemente XML.

Za večino izdaj v sklopu eZISS so bila uporabljena priporočila TEI P4 (Sperberg-McQueen, Burnard, 2002), med trajanjem projekta pa je bila izdelana nova inačica priporočil, TEI P5 (TEI, 2007), ki smo jo uporabili pri kasnejših izdajah (ŠP, Prisege).

Vsaka izdaja eZISS sestoji iz enega dokumenta XML (kar v večini primerov pomeni tudi eno datoteko, v primeru BS ali ŠP pa več deset datotek), ki je skladen s shemo TEI, prilagojeno za eZISS, znotraj nje pa uporablja

določen nabor elementov, ki pokrivajo specifične določene izdaje. V nadaljevanju izpostavimo nekaj bolj pomembnih vidikov takšnega kodiranja glede na posebnosti knjižnice eZISS. Kot primer zapisa izdaj eZISS podamo v Sliki 1 izseček eZISS izdaje romana Izidorja Cankarja *S poti*.

2.1.1 Kodni nabor

V vseh izdajah uporabljamo nabor znakov Unikod v kodiranju UTF-8. Ta univerzalni nabor znakov zadošča za večino (tudi historičnih) znakov, ki jih srečamo v posameznih izdajah, izjema so edino BS, kjer zaradi kompleksnosti uporabljenih znakov uporabljamo dodatni mehanizem, dostopen v TEI, s katerim je mogoče posebej označiti nestandardne znake in jim nato dodati standardno (vendar mogoče ne popolnoma ustrezno) interpretacijo po Unikodu ali pa nestandardno, vendar natančno interpretacijo s posebnimi tipografijami – pri BS smo uporabili pisavo ZRCola avtorja Petra Weissa. Spletna (HTML) izpeljanka BS je tako dostopna v dveh med seboj povezanih inačicah, kjer ena uporablja samo standardne znake Unikod, druga pa je vezana na pisavo ZRCola.

2.1.2 Kolofon TEI

Vsak dokument TEI vsebuje tudi kolofon (»TEI header«), ki opisuje označeno elektronsko besedilo, tako da so tekst sam, njegov vir, kodiranje in revizije temeljito dokumentirani. Kolofoni posameznih izdaj so lahko zelo obširni, saj vsebujejo tudi podatke, kot so definicije posebnih znakov (BS), podroben opis samih rokopisov (ŠP) itn.

2.1.3 Struktura besedila

Smernice TEI vsebujejo bogat nabor oznak za opisovanje strukture besedila, bodisi proznega (razdelek, odstavek, navedek ...), pesniškega (kitica, verz), dramskega (didaskalija, govorec, govor ...), slovarskega (geselski članek, iztočnica, definicija, primer ...) itn.

Edicije eZISS vsebujejo mdr. vse naštete besedilne zvrsti, zato uporabljajo zelo širok nabor strukturnih oznak.

2.1.4 Multimedijske vsebine

TEI nudi zelo dobro podporo za vključitev faksimilov v besedila, kjer je možno posamezni faksimile podrobno opisati in do njegove grafične podobe dostopati v več velikostih oz. resolucijah. To podporo v eZISS dodobra izkoristimo, saj skoraj vse edicije vsebujejo poleg prepisov tudi faksimile. Podpora za druge multimedijske vsebine je v TEI nekoliko slabša, zato smo za vključitev govora (BS) uporabili splošni mehanizem kazalk na avdio posnetke, medtem ko smo video (ŠP) shranili posebej, in povezavo med videom in besedili izvedli v okviru standarda Synchronized Multimedia Integration Language, SMIL.

2.1.5 Tekstnokritični aparat

Izdaje eZISS vsebujejo mdr. diplomatične prepise, ki ohranjajo izvorno besedilo skupaj z vsemi popravki in kasnejšimi dodatki ter spremembami naknadnih piscev oz. izdaj besedila. Za opis takšnih struktur nudi TEI posebne elemente za tekstnokritični aparat, ki lahko v besedilu izpostavijo dele, kjer je neko besedilo prečrtano ali dodano, kjer dodamo uredniške popravke ali razvezane kratice oz. kjer postavimo variantna branja (glej Sliko 1). Ti elementi so uporabljeni v skoraj vseh izdajah eZISS.

2.1.6 Navzkrižne povezave

Tipična edicija eZISS vsebuje besedilo kot faksimile in v več prepisih, ki so nato med seboj poravnani na nivoju strani, vrstice ali odstavka. Za realizacijo takšnih vzporednih prikazov smo uporabili metode navzkrižnih povezav, ki jih ponujajo smernice TEI, pri čemer smo zaradi specifik posameznih izdaj posegli po različnih metodah. Vsem metodam je skupno to, da vse elemente, ki naj bi bili poravnani, npr. vrstici diplomatičnega in

kritičnega prepisa, opremimo z enoznačnimi identifikatorji, ti pa so potem tako ali drugače medsebojno povezani.

2.2 Postavitev na splet in dostopnost

Če predstavlja dokument, zapisan v XML, »dejansko« digitalno izdajo, ki naj bi bila formalno pravilna, večnamenska in odporna na tehnološke spremembe, pa ta oblika ni najbolj primerna za branje ali iskanje po izdaji; za te namene je bistveno boljši HTML, saj je ta zapis primeren za neposredni prikaz v spletnih brskalnikih. Pretvorba iz TEI-XML v HTML je za vsako edicijo narejena avtomatsko, s pomočjo skript v jeziku XSLT. Za čim boljše predstavitev izdaj smo za vsako napisali namensko pretvorbo XSLT, katere rezultat čim bolje prikaže posebnosti izdaje v spletnem brskalniku. Kot primer je na Sliki 2 podan izgled izsečka romana *S poti*, ki ustreza izvornemu zapisu na Sliki 1.

Tako vsaka izdaja eZISS na spletu vsebuje naslednje komponente:

- krovna stran, s kratko predstavitevjo izdaje in povezave,
- izdaja v HTML, primerna za pregledovanje in branje v spletnih brskalnikih,
- kanonična izdaja v XML-TEI, skupaj s shemo XML in skripto XSLT za pretvorbo v HTML.

Vse izdaje je možno brati neposredno na spletu, dostopne pa so tudi kot stisnjene datoteke v formatu ZIP, tako da jih je možno (skupaj s faksimili in drugimi multimedijskimi vsebinami) prenesti na lastni računalnik in uporabljati brez spletne povezave. Za prenos je dostopen tako HTML kakor tudi izvorni XML-TEI, primeren za nadaljnje računalniške obdelave.

Vse izdaje so dostopne pod eno od licenc Creative Commons, kar pomeni, da jih je dovoljeno v celoti kopirati in razširjati. Pri vseh izdajah zahtevamo priznanje avtorstva, odvisno od izdaje pa tudi deljenje pod enakimi pogoji oz. pogoj, da se izdaja ne predeluje.

Dostop do spletnih strani knjižnice eZISS spremljamo s programom Google Analytics, ki nam pokaže, da strani v povprečju obiše okoli 50 uporabnikov na dan, od tega polovico iz Slovenije. Okoli polovica vseh uporabnikov najde knjižnico oz. posamezne izdaje prek spletnih iskalnikov (kjer sta najbolj pogosta iskalna niza »brižinski spomeniki« in »anton martin slomšek«), tretjina pa prek povezanih strani. Približno četrtnina obiskovalcev uporabi knjižnico večkrat, okoli 6 % obiskovalcev (v zadnjem letu okoli 1.000) pa se na straneh zadržuje več kot 10 minut.

3 Sklep

Knjižnica eZISS (<http://nl.ijs.si/e-zrc/>) se trudi narediti temeljna besedila slovenskega slovstva bralcem kar najširše dostopna in med humanisti (zlasti filoloških usmeritev) povečati zanimanje za raziskovanje primarnih virov slovenske literature. S tem je povezana skrb za natančen in standardiziran zapis posameznih izdaj, kot tudi zelo liberalni pogoji uporabe, s katerimi prednjači v Sloveniji, kakor tudi v širšem regionalnem prostoru. V izdajah eZISS so uporabljene različne metode analize in prepisovanja besedil (diplomatični, kritični, fonetični prepis), pa tudi kombinacija prepisov in drugih medijev (faksimile, zvočni in video posnetek), da bi tako omogočili raznolike poglede in pristope k besedilu.

Tehnološka podlaga izdaj se opira na prevladujoče, v svetu sprejete standarde elektronskega označevanja besedil – XML, TEI, Unikod –, kar izdajam zagotavlja odpornost na tehnološke spremembe. Vendar tudi znotraj teh tehnologij razvoj nenehno napreduje in bomo verjetno tudi naše starejše izdaje prilagodili najnovejšim tehnološkim izboljšavam (prehod iz TEI P4 v P5), načrtujemo pa tudi drugačne možnosti dostopa do edicij, predvsem prek iskalnega vmesnika. Izdaje projekta eZISS udeležujejo združitev klasičnega filološkega raziskovanja besedil z modernimi postopki za njihovo hranjenje, analizo in prezentacijo. Upati je, da

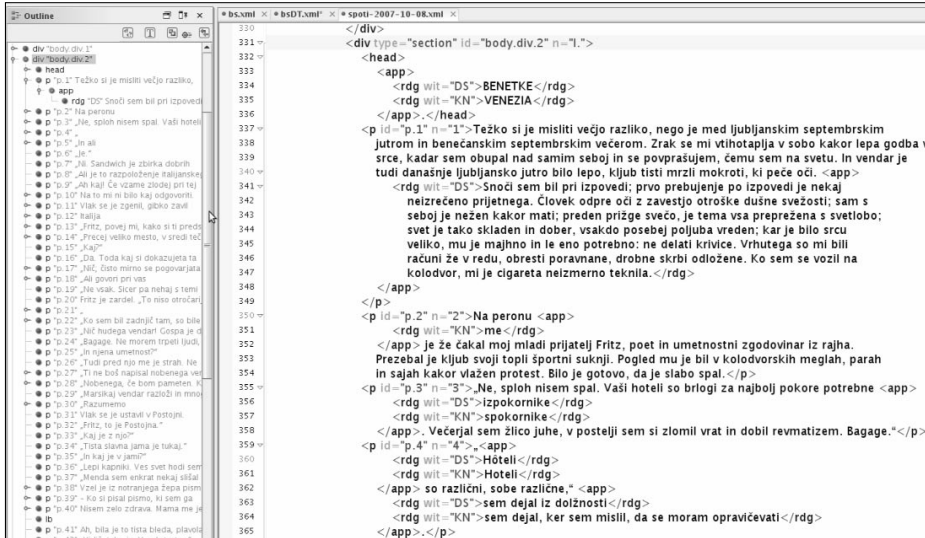
bo takšna združitev prispevala k oživitvi širšega zanimanja za raziskovanje primarnih virov slovenskega slovstva.

Zahvala

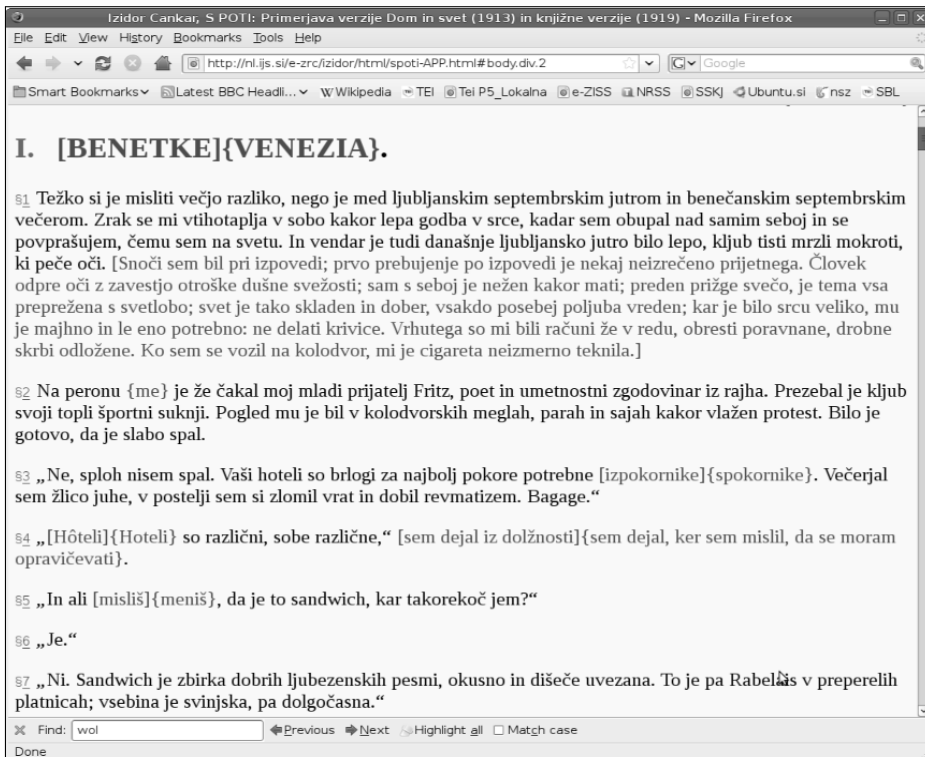
Delo, predstavljeno v tem prispevku, so finančno podprli Ministrstvo za visoko šolstvo, znanost in tehnologijo Republike Slovenije (MVZT), Agencija za raziskovalno dejavnost Republike Slovenije (ARRS) in Slovenska akademija znanosti in umetnosti (SAZU) v okviru aplikativnih projektov: *Znanstvenokritične izdaje v elektronskem mediju*, *Elektronske znanstvenokritične izdaje slovenskega slovstva* in ciljnega raziskovalnega projekta *Elektronsko besedilno središče*.

Literatura

- BERNIK, France, GRDINA, Igor idr. (ur.), 1992, 1993: *Brižinski spomeniki. Znanstvenokritična izdaja*. Ljubljana: SAZU, ZRC SAZU.
- BURNARD, Lou, SPERBERG-MCQUEEN, C. M. (ur.), 2002: *Guidelines for Electronic Text Encoding and Interchange*. Text Encoding Initiative Consortium, TEI P4, XML-compatible edition: <<http://www.tei-c.org/P4X/>>. (Dostop 1. 8. 2009.)
- OGRIN, Matja (ur.), 2005: *Znanstvene izdaje in elektronski medij*. Studia Litteraria, ZRC SAZU.
- TEI Consortium (ur.), 2007: *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 1.4.1. Last updated on July 1st 2009. TEI Consortium: <<http://www.tei-c.org/Guidelines/P5/>>. (Dostop 1. 8. 2009.)



Slika 1: Izseček XML-TEI zapisa romana *S poti*, kjer vidimo strukturne elemente in tekstnokritični aparat, ki izpostavi razlike med objavo v reviji *Dom in svet* («DS») in predelano knjižno izdajo («KN»).



Slika 2: Izseček spletne (HTML) oblike romana *S poti*, kjer vidimo, kako se izvorne oznake TEI, podane v Sliki 1, preslikajo v vizualno usmerjene oznake.