

Ogrin, Matija; Erjavec, Tomaž

Znanstvenokritične izdaje v elektronskem mediju slovenski in češki metodološki pristopi k posredovanju besedil

Sborník prací Filozofické fakulty brněnské univerzity. X, Řada literárněvědné slavistiky. 2006, vol. 55, iss. X9, pp. [247]-256

ISBN 80-210-3987-6

ISSN 1212-1509

Stable URL (handle): <https://hdl.handle.net/11222.digilib/102979>

Access Date: 17. 02. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

MATIJA OGRIN, TOMAŽ ERJAVEC
(Inštitut za slovensko literaturo in literarne vede ZRC SAZU, Ljubljana
Odsek za tehnologije znanja, Inštitut ‚Jožef Stefan‘, Ljubljana)

ZNANSTVENOKRITIČNE IZDAJE V ELEKTRONSKEM MEDIJU SLOVENSKI IN ČEŠKI METODOLOŠKI PRISTOPI K POSREDOVANJU BESEDIL

1 Splošne opombe

Metodologija znanstvenih izdaj je doživela z razmahom informacijske tehnologije ponoven razcvet. Elektronske znanstvene izdaje in viri so v razvitih družbah postali pomemben medij obdelave, posredovanja in prezentiranja informacij na področju humanističnih ved. Termin „Digital Humanities“ – digitalna humanistika – označuje široko dejavnost, ki vključuje raziskovanje, objavljanje in analiziranje jezikovnih in slovstvenih dokumentov. Pri tem se razvijajo in tudi spreminjajo metodološke osnove posredovanja in predstavljanja besedil, saj ima računalniški medij za to velike zmogljivosti. Ni pretirano reči, da je podpora informacijskih tehnologij humanističnim vedam marsikje oživila interes za historične plasti jezika, kulture in literature, interes za empirično faktografijo, ki se realizira v visokozmogljivih iskalnih orodjih elektronskih jezikovnih in besedilnih zbirk, v vzporejanju digitalnih faksimilov in transkripcij, v detajlizmu znanstvenega aparata.

Fundamentalno gledano se elektronske besedilne zbirke (če mislimo na njihov objekt) razlikujejo po tem, za kakšen način posredovanja besedilnega izročila se odločajo. Teoretično sta na voljo dve skrajni možnosti: na eni strani „knjižnica“ kot zbirka knjig, v katerih uredniki podajajo svoje videnje, svojo redakcijo (in svojo selekcijo) besedil. Druga skrajna možnost je „arhiv“ originalnih rokopisov, ki se, povsem nedotaknjeni od zunanjih interpretativnih posegov, bralcu kažejo v svoji izvorni zgodovinski prvobitnosti, pa tudi drugačnosti in nerazumljivosti. Prva možnost, koncept „knjižnice“, nam posreduje besedilno tradicijo v izborih in prepisih na vsaj deloma interpretiran način. Druga možnost, koncept „arhiva“, nam tradicijo posreduje v zavestni abstinenci vsake ambicije po interpretiranju besedila. Ambicija „knjižnice“ je garantirati razumevanje besedil: dospeti na konec komunikacijske verige; v središču tega koncepta je bralec, točneje, bralčeva nemotena recepcija tradicije. Ambicija „arhiva“ je, nasprotno, najti kontakt z originalno tradicijo in jo ohranjati, brez edicijske intervencije, četudi za to ceno, da bo besedilno izročilo ostalo (vsaj deloma) nerazumljivo zaradi starih form jezika, starih pisav, pisarskih konvencij itn.; v središču koncepta „arhiv“ je torej – lahko bi rekli – prvi člen komunikacijske verige in začetek tradicije: originalni rokopis, avtograf.

Koncept „knižnice“ ima nepogrešljivo kulturno vlogo, saj tradicijo posreduje naprej in zagotavlja kulturno kontinuiteto. Vendar ima lahko šibko točko, da v nekaterih primerih besedilno izročilo prispe do bralca v obliki, ki je preveč modificirana. Bralčev stik z izročilom je v teh primerih nepopoln v „kvalitativni“ dimenziji. Koncept „arhiva“ ima, nasprotno, vlogo ohranjanja, konzerviranja tradicije; toda ker bo ta ostala v starih jeziki in pisavah bolj ali manj nerazumljiva, bo stik z njo zelo pomanjkljiv v „kvantitativni“ dimenziji.

Elektronski medij omogoča tako prvo kot drugo možnost posredovanja besedil – omogoča pa tudi določeno kombinacijo in notranjo uskladitev obeh principov; zdi se, da je to ena njegovih očitnih prednosti.

2 Trije pristopi

Kar zadeva trenutno raznolikost pristopov k posredovanju besedil pri Čehih in Slovencih, je mogoče konstatirati, da imajo na tem področju primerljive ali podobne izkušnje; vendar je očitna prednost Češke v tem, da nekatere njene institucije že več let sistematično digitalizirajo nekatere svoje zbirke v velikem obsegu. To velja zlasti v dveh primerih.

Prva takšna institucija je Národní knihovna ČR, oddelek za rokopise in stare tiske, ki vodi projekt Memoria, znotraj njega pa nastaja zelo velika zbirka digitaliziranih rokopisov *Manuscriptorium* (http://www.memoria.cz/site_cz/). Že doslej so v tem projektu skenirali več kot milijon strani dragocenih rokopisov in redkih starih tiskov; delo poteka kot koordinirano prizadevanje številnih čeških, pa tudi slovaških idr. knjižnic. Skenogrami so posneti v visoki ločljivosti in omogočajo podroben pregled vsakega folija rokopisa. Vsaka rokopisna knjiga oziroma enota je v digitalni zbirki *Manuscriptorium* predstavljena s po enim kvalitetnim skenogramom za vsako stran in s podrobnim metapodatkovnim kodikološkim opisom. Ti metapodatki so strukturirani po shemi Manuscript Access through Standards for Electronic Records (MASTER – <http://xml.coverpages.org/master.html>), ki je nastajal v tesni povezanosti s konzorcijem Text Encoding Initiative. Vendar so v Češki nacionalni knjižnici to shemo za svoj projekt modificirali z nekaterimi elementi, mdr. z elementi za prepis nekaterih delov rokopisa in z elementi za definicijo barv digitalnih posnetkov; svojo shemo so poimenovali MASTER+. Zbirka je opremljena z iskalnikom, iskati je mogoče po številnih kategorijah iz te razširjene sheme, pogled na rokopise pa je prilagodljiv različnim osebnim preferencam. Javnosti so na voljo le metapodatki in slike strani v nizki ločljivosti, ki ne omogoča branja, medtem ko je za dostop do skenogramov visoke ločljivosti potrebna plačljiva licenca. Kljub tej omejitvi ponuja digitalni korpus *Manuscriptorium* izjemno bogastvo evropske rokopisne dediščine srednjega in zgodnjega novega veka. V *Manuscriptoriumu* je mogoče, kakor pove že ime, dostopati do impozantnega števila slik originalnih dokumentov nadvse dragocene pisne in likovne tradicije, ki tako prihaja pred bralca v svoji prvobitni naravi, brez posegov razlage in transkripcije (v novejšem času poteka tudi okrepljeno prepisovanje manjšega dela rokopisov v normaliziranih prepisih, ki še niso na voljo javnosti). Namen tega velikega dela je nedvomno posredovati javnosti podobo izvirnika in zavarovati originalni kodeks. V središču tega impozantnega podjetja bi lahko videli rokopis, kodeks kot izvir besedilnega izročila, ki ga želi projekt ohraniti in brez uredniških interpolacij odpreti javnosti. Za branje teh besedil je kajpada potrebno obsežno in specializirano znanje: najprej zgodovinsko in kodikološko znanje, da iz zunanosti kodeksa razberemo znamenja, ki ga umeščajo v čas

in prostor; nato nastopijo srednjeveški jeziki in dialekti, molčeče pisave in nerazumljiva znamenja – vse te prvine spregovorijo le filološko in paleografsko izurjenemu človeku. Razumevanje teh besedil je torej mogoče le znanstvenikom, ki pripravljajo kodikološki opis v omenjeni shemi metapodatkov. V tem smislu ustreza *Manuscriptorium* konceptu odprtega „arhiva“, ki brez razlage čaka na razumevanje poznavalca – ali da iz njega kdo pripravi izdajo kot del „knjižnice“.

Druga institucija, ki gradi obsežen elektronski vir, je Ústav pro českou literaturu AV ČR, ki je leta 2002 izdal veliko zbirko češke poezije 19. stoletja na CD-Romu. (Svatbová 2002) Do leta 2005 pa je tej ustanovi zunanji izvajalec celoten korpus poezije pretvoril v novejšo tehnologijo (XML) in je zdaj objavljen na svetovnem spletu (<http://www.ceska-poezie.cz/cek/>). Korpus *Česká elektronická knihovna* omogoča brezplačen dostop do besedil 1200 čeških pesniških zbirk v diplomatičnem in kritičnem prepisu z možnostjo vzporedne prezentacije obeh prepisov; ne vključuje digitalnih posnetkov originalnih tiskov, vključuje pa obsežno uredniško in metapodatkovno gradivo z bogatim opisom prvotiska, bibliografskimi podatki itn. Posebna kvaliteta tega odličnega elektronskega vira je to, da je po njem mogoče iskati z raznovrstnimi orodji: iskati je mogoče poljubne besede ali zveze na ravni zbirk(e), pesmi, verza in predstaviti rezultate v kontekstu kot konkordanco, povezano z glavnim besedilom; iskati je mogoče vzorce pod-besednih enot, kakor so zlogi in rime; elektronska orodja lahko bralcu za poljuben izbor pesniških zbirk pripravijo frekvenčni slovar njihovega besedja ali abecedni slovar s frekvencami itn. *Česká elektronická knihovna* torej izpričuje večplastno literarnovedno analitično koncepcijo, ki je dosledno in praktično uresničena s širokim naborom iskalnih orodij. Vse to, skupaj z obsežnimi uredniškimi podatki, odpira številne možnosti tako za panoramski pregled nad razvojem češke poezije kakor za minuciozno literarnoteoretsko analizo – morfološko, poetološko, verzološko idr. Glede na ta koncept, ki daje prednost recepciji in najrazličnejšim stopnjam razumevanja češke poezije, bi *Česká elektronická knihovna* ustrezala predvsem konceptu „knjižnice“. Hkrati je ta odličen polnobesedilni korpus zavarovan pred odstopanji od originalne tekstualne tradicije s tem, da poleg uredniškega, t. j. kritičnega prepisa vseh pesmi vključuje tudi diplomatični, torej literalni prepis vse poezije. Tako je videti, da *Česká elektronická knihovna* udejanja eno najboljših možnih izbir, saj v polni meri izpolnjuje poslanstvo „knjižnice“ – posredovanje, široko recepcijo, razumevanje izročila – hkrati pa s principi diplomatičnega prepisovanja zagotavlja tej tradiciji pristnost.

V Sloveniji ni še nobena primerljiva ustanova začela s tako obsežno zastavljeno digitalizacijo katerega koli gradiva, ki bi ga nameravala tudi sistematično urediti kot elektronski vir z znanstveno zasnovanim konceptom. Pač pa je v Sloveniji prve korake v to smer napravil projekt *Elektronske znanstvenokritične izdaje slovenskega slovstva* (<http://nl.ijs.si/e-zrc/>), ki na Inštitutu za slovensko literaturo in literarne vede ZRC SAZU in s sodelovanjem Odseka za tehnologije znanja IJS poteka od leta 2001. Ta projekt poteka v neprimerljivo manjšem merilu, na gradivu, ki je v primerjavi z obema češkima elektronskima viroma miniaturno. Vendar imajo *Elektronske znanstvenokritične izdaje slovenskega slovstva* nekatere posebnosti. Te izhajajo iz naslednjih temeljnih izhodišč.

- Eden glavnih namenov našega projekta je bil premisliti vodila tradicionalne ekdotike, jih aplicirati na izdaje v elektronskem mediju in v kontekstu elektronske izdaje ta pravila ponovno artikulirati – tako glede na možnosti elektronskega medija kakor glede na zahteve, ki jih postavlja specifično gradivo slovenskega slovstva v njegovi zgodovinski in literarnozvrstni raznolikosti.
- Zasnovati elektronske izdaje na odprtih mednarodnih standardih elektronskega kodiranja oz.

označevanja, ki zagotavljajo ne le raznovrstno uporabnost označenih besedil ter mnogotere možnosti njihovega prikaza in analize, ampak tudi njihovo neodvisnost od tehnoloških sprememb, trajnost ter izmenljivost.

- Pomemben motiv je bil tudi nadaljevati tradicijo znanstvenih izdaj našega Inštituta, znotraj tega pa želja, posredovati dragocena, a manj znana besedila slovenskega slovstva širši kulturni ter znanstveni javnosti v prostem dostopu na svetovnem spletu.

Iz teh zasnov so se razvile izdaje, ki poskušajo prikazati kompleksno naravo objavljenih besedil, omogočiti različne poglede ali dostope k tem besedilom in s tem odgovoriti kar se da raznolikim raziskovalnim interesom.

3 Primer Treh pridig o jeziku

Prva izdaja, ki je nastala v okviru tega projekta in udejanja poglavitne naštetih cilje, so *Tri pridige o jeziku* (<http://nl.ijs.si/e-zrc/slomsek/>) Antona Martina Slomška, trije reprezentativni primeri retorske proze, ki je v 19. stol. doživela vrh razvoja prav s Slomškom. Takšna besedila so možen vir različnih raziskav – o zgodovini slovenskega jezika, o literarnozvrstnih aspektih, o zgodovini retorike, možni so razni teološki pristopi itn. Zato je bilo osnovno vodilo pri tekstnokritični obdelavi vseh teh besedil, da mora izdaja tako večstranskega in večplastnega vira omogočiti vse možne načine njegovega preučevanja – v eni sami izdaji. Vsak vidik besedila, ki more postati predmet te ali one znanstvene obravnave, mora biti v izdaji dostopen vsaj v enem od prikazov besedila.

To izhodišče je zasnoval Jože Faganel (Faganel 2004) in ga uresničil tako, da je ločeval med tremi pojavnimi oblikami besedila. To so:

1. faksimile rokopisa,
2. diplomatični prepis in
3. kritični prepis.

Faksimile je vključen v izdajo kot digitalna fotografija srednjevisoke ločljivosti, ki omogoča veliko povečavo za podroben pregled rokopisa; koristen je zlasti za neposreden stik z avtografom, za evidenco nad avtorjevimi korekcijami, zanimivostmi pisave, položaja pomenljivih odlomkov, kot tudi za proučevanje retorične strukture besedil.

Diplomatični prepis je podrobno izdelan tipografski dvojnik rokopisa v vsemi avtorjevimi nedoslednostmi, napakami, avtorjevimi popravki, vrivki ipd. vred. Njegova funkcija je, da predstavlja izvornik in tako olajša branje težje berljivih mest rokopisa. Diplomatični prepis pušča kratice, nejasna in poškodovana mesta rokopisa nerazrešena in nedopolnjena; v našem primeru sta dve takšni poškodovani mesti v drugi pridigi. Diplomatični prepis ohranja tudi možnost branja prečrtanih delov besedila, s čimer se pokaže avtorjevo izboljševanje in redigiranje, vključuje pa tudi podčrtavanje besedila, ki je bilo verjetno opravljeno po končanem pisanju med učenjem pridige za govorni nastop in predstavlja retorične poudarke.

Kritični prepis je že tekstnokritična interpretacija besedila, ki izvorno besedilo modifikira na črkopisni in oblikoslovnih ravni, vendar se pri tem ravna po ekspliciranih načelih, pridobljenih z indukcijo iz analize samega besedila – upoštevajoč pri tem zvrstnost besedila, čas njegovega nastanka in namembnost znanstvene izdaje. Biblične reference so v kritičnem prepisu verificirane in s hiperbesedilnimi kazalkami povezane z elektronsko zbirko bibličnih tekstov na straneh Svetopisemske družne Slovenije (<http://www.biblija.net/>). Kritični prepis vsebuje tudi ustrezen aparat z opombami.

Vzporedni prikaz obeh prepisov je najbolj očitna prednost, ki jo omogoča elektronska izdaja. Na ravni posameznih vrstic sta paralelno prikazana oba prepisa istega teksta. Tak prikaz je uporaben ne samo za raziskovalne namene, ampak z ustrežno razlago tudi za pedagoško rabo, saj tovrstna predstavitev daje jasno sliko o zgodovinskem razvoju slovenskega jezika.

Ker je predstavitev kompleksne narave besedil neposredno odvisna od zmožnosti ustrezne rabe odprtih standardov računalniškega označevanja, bomo v nadaljevanju več pozornosti posvetili označevalnim metodam, ki jih uporabljamo v našem projektu.

4 Zapis izdaj

Besedila, ki naj imajo takšno kompleksno strukturo, je potrebno primerno računalniško obdelati. Besedilno strukturo (različne besedilne enote) in znotrajbesedilne relacije (formalne in pomenske odnose med deli besedila) – je potrebno eksplicirati tako, da postanejo za računalnik „razumljive“ na nedvoumen način, s tem pa manipulabilne in lahko tako z našo izdajo dosežemo določene namene. To dosežemo s tako imenovanim kodiranjem ali označevanjem besedil. Potrebujemo torej opisni računalniški meta-jezik, s katerim računalniku „povemo“, kaj posamezni deli besedila predstavljajo in kako se med sabo povezujejo. Jezik, ki omogoča razmeroma razumljiv, a dovolj kompleksen opis besedilne strukture, se imenuje eXtensible Markup Language – XML. Med njegovimi mnogimi prednostmi je pomembna za označevanje besedil zlasti ta, da je vprašanje prikaza (kako bo besedilo videti na zaslonu) povsem izločeno, pač pa je v osredje postavljeno, kaj določen del besedila pomeni v strukturi celotnega besedila in na kaj se navezuje. Tako denimo niz `<l>Byl pozdñi večer – prvni máj – </l>` računalniku nedvoumno pove, da vsebovano besedilo predstavlja verz, in če bi začetno oznako zapisali kot `<l n='l'>`, bi računalniku s tem povedali, da gre za prvi verz pesmi in bi na tej podlagi v večji besedilni zbirki že lahko iskali po kriteriju prvih vrstic ali iz njih strojno generirali kazalo prvih verzov ipd.

Pravila, kateri elementi smejo vsebovati druge elemente, kakšna so njihova razmerja, kdaj smejo nastopati in kdaj ne, so zbrana v neki zunanji zbirki pravil, ki se imenuje definicija tipa dokumenta (Document Type Definition) ali DTD. Na začetku dokumenta XML je povedano, v kateri datoteki DTD so ta pravila podana in kje je shranjena. S tem so računalniku nedvoumno podana pravila za uporabo besedilnih elementov, kar je vitalna lastnost jezika XML, saj lahko programsko preverimo, ali je struktura našega besedila skladna s pravili, določenimi v DTD. Če je tako in je torej datoteka XML veljavna, to omogoča vrsto nadaljnjih korakov, v primeru da ni, pa razčlenjevalnik sporoči mesto in vrsto napake, kar olajša popravljanje dokumenta.

Teoretično bi si lahko zaželeno shemo pravil – torej DTD – izdelali sami glede na posebne potrebne naše izdaje. Toda ker to ni preprosta naloga, predvsem pa zato, ker bi naše elektronsko besedilo v tem primeru ne bilo kompatibilno ter izmenljivo z nobeno drugo besedilno zbirko, tega ni vredno poskusiti, vsaj ne za kompleksne in izmenljive izdaje. Rezultati so mnogo boljši, če sprejmemo že obstoječo označevalno shemo konzorcija TEI.

Smernice za označevanje besedil Text Encoding Initiative Guidelines (Sperberg-McQueen & Burnard, 2002) so nastajale kot konsenz široke akademske skupnosti vse od konca osemdesetih let. Danes predstavljajo odprt de facto standard besedilnega označevanja in objavljanja. Te smernice ponujajo širok spekter oznak, prilagojen tako velikim enotam kakor potankostim besedilnih in jezikovnih struktur. Definirajo tudi t. i. glavo, to

je razčlenjen kolofon izdaje, kamor urednik vpiše metapodatke o izdaji na način, skladen z mednarodnimi bibliotekarskimi standardi (MARC). Smernice TEI določajo tudi način, kako naj avtorji izdaje dokumentirajo svoja specifična odstopanja od predpisane sheme; smernice zahtevajo torej standardiziran opis morebitnih nestandardnih dodatkov k označevalni shemi.

Zaradi tendence po celostnem zajemu besedilnih struktur v smernicah TEI ni določena le ena shema, ki bi obsegala vse vrste besedil. Sedanja verzija priporoči (TEI P4, <http://www.tei-c.org/P4X/>) namesto tega vsebuje vrsto modulov, ki jih je mogoče poljubno izbrati in s tako prilagojeno kombinacijo dobiti želeni nabor dovoljenih elementov in njihovih razmerij, torej želeni DTD – definicijo tipa dokumentov v XML. Za projekt Elektronskih znanstvenokritičnih izdaj slovenskega slovstva smo izbrali tele module:

- TEI.prose, osnovni modul za označevanje proznih besedil – vsebuje elemente za glavo TEI (metapodatki, ki opisujejo datoteko, vir besedila, izdajo, uredniško prakso itn.), kakor tudi standardne elemente za strukturiranje besedila (razdelek, odstavek, opomba, tabela idr.) in nekatere za pod-odstavčno označevanje (raznovrstni poudarki);
- TEI.transcr, dodatni modul za prepis primarnih virov, zlasti rokopisov, ki vsebuje elemente za korekcije in emendacije, označevanje različnih pisarskih rok in številnih drugih rokopisnih posebnosti;
- TEI.linking, dodatni modul za znotraj – in medbesedilne povezave z elementi in atributi, ki povezujejo različne verzije besedila in besedilo povezujejo z zunanjimi podatkovnimi viri;
- TEI.figures, dodatni modul za označevanje slik in grafičnega gradiva, uporabljen za označevanje faksimila; gre za predpisane oblike kazalk h grafičnim datotekam, ki vsebujejo skenogramme izvornika v različnih velikostih in resolucijah;
- TEI.extensions, specifičen modul za uporabnika, kjer so implementirane uporabnikove razširitve standarda TEI; uporabili smo ga za določitev drugačne vrednosti nekaterih atributov, denimo za mesto, kjer se nahaja informacija o opombah.

5 Priprava gradiva

Potek priprave elektronskih izdaj je v našem projektu običajno sledil v treh fazah:

1. priprava besedila v urejevalniku besedil (Word),
2. konverzija besedila v semantično bogaten zapis XML/TEI ali „pretvorba navzgor“ (up-conversion); ta oblika predstavlja „kanonično“ obliko naših izdaj,
3. konverzija kanoničnega besedila v zapis HTML ali „pretvorba navzdol“ (down-conversion); v zapisu HTML je gradivo primerno za branje in pregledovanje na zaslonu, sam zapis pa je pomensko osiromašen, saj so oznake v HTML bistveno bolj usmerjene na zunanji videz dokumenta, ne pa na njegovo interpretacijo.

V nadaljevanju poskušamo opisati te razvojne stopnje in tehnologijo, uporabljeno za analizo in predstavitev besedil.

V primeru naših elektronskih izdaj je izčrpna tekstokritična oz. ekdotična priprava teksta potekala v urejevalniku besedil (običajno Word). Že med tem delom je potrebno misliti na to, katere strukturne lastnosti besedila želimo označiti, t. j. eksplicirati z oznakami TEI, kar pomeni, da vložimo v besedilo določene informacije na tak način, da bo program iz besedila lahko pridobil potrebne dodatne informacije za njegovo obdelavo, prikaz ipd. (Hockey, 2000) Ko so bili diplomatični in kritični prepisi pripravljeni in opremljeni z opombami ter komentarji, je bilo gradivo pretvorjeno v kanonični zapis XML/

TEI. Ta obogatitvena konverzija ali pretvorba navzgor iz formata Word v XML je bila izpeljana najprej z enim od mnogih pretvornikov, kakor denimo OpenOffice in UpCast, v grobi zapis XML. Nato so bili za vsako od naših edicij pripravljeni specifični transformacijski programi. Ti pretvorijo grobo datoteko XML v več zaporednih korakih v ciljni zapis TEI. Ti programi ali „filtri“ so večinoma napisani v jeziku XSLT – transformacijskem jeziku za dokumente XML, ki je prav tako standardizirana specifikacija konzorcija W3C in jo zato podpirajo številna orodja, denimo novejši brskalniki (Firefox, IE). Čeprav je XSLT idealen za zapis strukturnih pretvorb dokumentov XML, pa je manj primeren tam, kjer naj bi strukturo XML šele pridobili iz različnih vzorcev, ki se skrivajo v samem besedilu. Tovrstni filtri so bili napisani v jeziku Perl. Naposled smo napisali tudi ustrezno stilno datoteko XSLT, primerno za prikaz dokumenta na zaslonu v obliki HTML.

Izdelava končne oblike izdaje ima obliko cikličnega procesa. Ko je pripravljena prva verzija gradiva in transformacijskega programja, pridobljeno gradivo v XML/TEI pretvorimo v HTML in ovrednotimo rezultate. Pokazalo se je, da so napake, ki jih pri tem lahko odkrijemo, treh vrst: 1. napake v originalni datoteki Word; 2. napake v transformacijskih scenarijih; 3. nepopolnosti v konceptu samega označevanja. V prvem primeru je potrebno popraviti izhodiščno besedilno datoteko, v drugem prilagoditi transformacijske datoteke XSLT, v tretjem pa je potrebno poseči v semantiko sheme XML (DTD), ki določa nabor uporabljenih elementov. Ko so te napake odpravljene, se pretvorbo navzgor ponovno požene in tako se celoten cikel po potrebi večkrat ponovi z različnimi dopolnitvami – bodisi vsebinske ali formalne narave.

V tem procesu gradivo kmalu doseže določeno mejo, ko dodatkov in popravkov strukture ni več mogoče racionalno vnašati in izvorno datoteko (Word), saj so možnosti za strukturiranje gradiva tu zelo omejene. Na tej stopnji Wordovo datoteko zavžemo in obdržimo le kanonično datoteko v zapisu TEI. Vsa dopolnila nato vnašamo neposredno v to kanonično verzijo v urejevalniku za XML. Od tod naprej je potrebno, da oseba, ki ureja besedilo, razume koncepte jezika XML in označevalno shemo TEI. Na tej stopnji priprave napišemo tudi metapodatke o izdaji, torej opišemo gradivo, njegov vir, način označevanja, historiat dopolnjevanja izdaje itn. – informacije, zbrane v glavi elektronske izdaje.

6 Predstavitev gradiva

Ko je gradivo urejeno in kodirano v kanoničnem zapisu, primernem tudi za shranjevanje, smo naposled še pred vprašanjem, kako gradivo čim bolj uporabiti, se pravi, kako ga objaviti na spletu ali na CD-ROMu.

Trenutno podpiramo za vsako knjigo v projektu eno pretvorbo v obliko HTML; ta je narejena s slogovno pretvorno datoteko XSLT še pred postavitvijo na splet. Naš HTML je razmeroma kompleksen, saj skuša prikaz čimbolj slediti izvirnemu rokopisnemu gradivu v vseh podrobnostih, kakor so npr. struktura besedila, njegov položaj na foliju, poudarki, avtorjeve korekcije, uredniške emendacije; naša predstavitev gradiva v HTML ponuja tudi vzporedni prikaz več besedilnih form, kakor so digitalni faksimile rokopisa, diplomatični prepis, kritični prepis ipd., kot tudi nekatere dodatke, možne le v digitalni izdaji, kakor je denimo povezava bibličnih citatov s kazalkami na informacijski vir Biblija.net, kjer bralec lahko preveri natanko določeno vrstico besedila v raznih, tudi historičnih slovenskih prevodih Svetega pisma. Pri zunanji podobi naših elektronskih besedil smo bili pozorni tudi na standarde o splošni dostopnosti gradiva osebam z motnjami vida (Wymer 2005).

Pretvorba navzdol v HTML je narejena posebej za vsako izdajo, čeprav so nekateri deli teh programov podobni ali enaki. Poleg grafične realizacije nekaterih elementov izvorne datoteke v zapisu TEI (denimo pretvorba TEI elementov v HTML elemente <s> za ‚strikethrough‘, prečrtano) mora pretvorba navzdol generirati kazalo vsebine, in kar je bistvenega pomena, prikaze s poravnavo digitalnih faksimilov in besedila, kakor tudi vzporeden prikaz prepisov.

Za raziskovalca, ki si želi izvedeti kaj več o kateri od elektronskih izdaj, je v prikazu HTML pomembna tudi glava TEI; pri konverziji le-te se namreč elementi TEI (kot je npr. <publicationStmnt>) izpišejo opisno, v slovenskem jeziku (Izjava o objavi), poleg tega pa so elementi povezani s svojimi definicijami v priporočilih TEI. Ker glava vsebuje tudi seznam vseh elementov, uporabljenih v izdaji, je tako neposredno dostopna dokumentacija za vse elemente TEI, uporabljene v posamezni elektronski izdaji.

7 Sklep

Postopki elektronskega kodiranja so tehnična zrcalna podoba ekdotičnih konceptov in postopkov. Kjer je struktura besedil bolj kompleksna in kjer izdaja vsebuje več analitičnih orodij, tam je v ozadju več elektronskega označevanja in programiranja. V tem pogledu je najbolj kompleksno razčlenjena elektronska knjižnica češke poezije. Vendar obstajajo pomembne razlike na ravni integriranosti ekdotičnih in računalniških postopkov znotraj posameznega projekta, pa tudi razlike glede tega, koliko posamezen elektronski vir temelji na mednarodnih odprtih standardih kodiranja, ki edini zagotavljajo relativno trajnost elektronskih besedil. Češki projekt poezije 19. stoletja se je namreč začel kot aplikacija nestandardnih Microsoftovih programskih orodij; ta pot ni omogočala trajne uporabnosti zbirke, zato so v letu 2003 začeli s konverzijo celotnega korpusa poezije v XML. Vendar je potrebno tehnološko znanje ostalo vezano na zunanjega izvajalca projekta, označevalna shema pa je neznana, skrita za spletni portal te elektronske knjižnice – kar ne prispeva k večji varnosti in trajnosti tega kvalitetnega elektronskega vira. Projekt *Manuscriptorium* pa je standardno označevalno shemo MASTER spremenil in razširil v MASTER+. Zdaj, ko je shema MASTER sprejeta v osnutek nove izdaje smernic konzorcija TEI kot relevantna shema za opis rokopisov, se odpira vprašanje, koliko je in bo opis rokopisov velikega korpusa *Manuscriptorium* skladen s priporočili TEI, kar lahko vpliva na stopnjo izmeljivosti teh podatkov.

Kar zadeva podrobno, ne-ambivalentno kodiranje tekstov, ki naj po eni strani integrira več možnih form besedila oziroma več pogledov nanj, po drugi strani pa naj z naslonitvijo na odprte standarde doseže čim večjo trajnost in izmenljivost – v tem pogledu pa imajo elektronske izdaje slovenskega projekta nekatere vsaj konceptualne, na manjšem gradivu realizirane prednosti. Te e-izdaje razmeroma podrobno dokumentirajo ne le razmerja med rokopisom in njegovimi prepisi, ampak tudi proces avtorjevih korekcij in uredniških emendacij. Hkrati je tu celotno označevanje javno dostopno (open source); celoten arhiv izdaje vključno s faksimili si lahko uporabnik prekopira in ga uporablja lokalno.

Prednost obeh čeških projektov je velika kvantiteta gradiva in uporabniška orodja, ki dragoceno gradivo napravijo dostopno množici uporabnikov. Vendar se zdi, da sta si v metodološkem pogledu kot ‚arhiv‘ (*Manuscriptorium*) in ‚knjižnica‘ (*Česká elektronická knihovna*) koncepta obeh velikih zbirk komplementarna. Tovrstna komplementarna združitev na tako obsežnem gradivu seveda ni preprosta; poleg tega velja upoštevati, da

koncepta obeh elektronskih virov izhajata iz narave ter poslanstva njunih ustanov. Kljub temu pa slovenski projekt *Elektronske znanstvenokritične izdaje slovenskega slovstva* napeljuje k misli, da je komplementaren pristop ne le možen, ampak je za znanstvene izdaje tudi koristen, kar potrjuje podobna praksa številnih zahodnoevropskih projektov.

Ena ključnih potez e-izdaj slovenskega projekta je namreč ta, da brišejo meje med arhivom, tradicionalno izdajo in knjižnico. (Sahle, 2005) Izdaja intergira v celovito predstavitev več prvin: izvornega nosilca besedila – rokopis ali prvotisk – pristno besedilo v diplomatičnem prepisu in kritično rekonstruirane, torej že interpretirane prepise. Da bi izdaja dosegla ta cilj, se morajo njeni ustvarjalci tudi sami vsaj do določene stopnje spoprijeti s sodobno informacijsko tehnologijo, ki v obliki „digitalne humanistike“ postaja tako rekoč ena od novih pomožnih ved humanistike.

Multifunkcionalnost in večplastna narava *Elektronskih znanstvenokritičnih izdaj slovenskega slovstva* naj bi – vsaj načeloma – zagotavljala, da je rokopis kot dragocena arhivalija varno shranjen in zavarovan pred nepotrebno uporabo v arhivu, medtem ko je potencial besedila, katerega rokopis nosi, odprt branju in razumevanju v raznovrstnih prepisih.

Rokopisu, shranjenemu v arhivu, tako ostane predvsem muzealska vrednost. Njegova kulturna vrednost in potencialnost pa lahko preideta v celostno koncipirano elektronsko izdajo, odtod pa v čim širše sfere znanosti, kulture in omike.


LITERATURA IN VIRI

- FAGANEL, Jože (2004): Osnovni podatki, opredelitev značaja izbranih besedil in raziskovalni interes. V: A. M. Slomšek: *Tri pridige o jeziku*. (<http://nl.ijs.si/e-zrc/slomsek/data/eSlomsek.html#slu1>)
- HOCKEY, S. (2000). *Electronic Texts in the Humanities. Principles and Practice*. Oxford University Press.
- SAHLE, Patrick (2005): *Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung*. http://www.germanistik.ch/publikation.php?id=Digitales_Archiv_und_digitale_Edition (9. januar 2006)
- SPERBERG-MCQUEEN, C. M., BURNARD, L. (eds.) (2002). *Text Encoding Initiative: Guidelines for Electronic Text Encoding and Interchange, TEI P4, the XML-compatible edition*. TEI Consortium. <http://www.tei-c.org/P4X/>
- SVATBOVÁ, Blanka, ur. (2002): *Poezie 19. století od thámovců po lumírovce. Ústav pro českou literaturu AV ČR, Praha, 2002.*
- WYMER, K. (2005). *Why Universal Accessibility Should Matter to the Digital Medievalist*. *Digital Medievalist* 1(1). <http://www.digitalmedievalist.org/>

Slika 1: Primer vzporednega izpisa

TRI PRIDIGE O JFZRIJL. Elektronska znanstveni inšna izdaja v zapisku XML - TTI (2004-10-25) L - Microsoft Internet Explorer

Address http://t1-01.aic.eurjformackfdata/cd/temack/v11.html#1a



	1825. XIII	
	1825. XIII	1.1 op. [1]
Na 16 nedelo po Binkulstih.		
Na 16. nedeljo po Binkulstih.		1.2 op. [2]
K'kerfhanikimu govorjenju		
K krščanskemu govorjenju		
nagovor.		
nagovor		1.4 op. -2]
Takrat bude tebi zheft, kader tjišti, kateri je tebe po=		1.5
Takrat bude tebi čest, kader tjišti, kateri je tebe po=		1.5
vabil, tebi porezhe: Prijatelj, pomekni le gori.	Luk. 14.	
vabil, tebi poreče: Prijatelj, pomekni se gori.	Luk. 14. [10]	
Vvod.		
Uvod		
1. Tolko štánov je na le temu zhašnimu fvetu, pa vñih le		
1. Tolko stánov je na le tem časnem svetu, pa vseh le		
eden pokliz: dofezhi nebesško kraletvo.		
eden poklic: doseči nebesško kraljjestvo.		
Od fvetiga zejarja na fedeshu fhatim do berazhe		1.10
roma=		
Od svetlega cesarja na sedežu zlatem do sroma=		1.10
ka per palzi berahki je vñaki fhan Bog is'volil		

Internet