

Trajnost digitalnih izdaj Uporaba statičnih spletnih strani na portalu Zgodovina Slovenije - Sistory

Andrej Pančur*

* Inštitut za novejšo zgodovino
Kongresni trg 1, 1000 Ljubljana
andrej.pancur@inz.si

Povzetek

Prispevek izhaja iz stališča, da je pri digitalnih izdajah potrebno poskrbeti za čim bolj celovito digitalno trajnost tako podatkov kot prezentacij, funkcionalnosti in programske kode. To je velik izziv predvsem za manjše digitalno humanistične projekte z omejenim financiranjem, ki ne omogoča dolgoročnega vzdrževanja tehnično zahtevnih digitalnih izdaj. Kot alternativno rešitev so v prispevku predstavljene rešitve, ki jih v zadnjih letih ponuja hiter razvoj statičnih spletnih strani. Digitalne izdaje, ki temeljijo na TEI, so s pomočjo osnovnih XML (XSLT) in spletnih tehnologij (HTML, CSS, JavaScript) kot statične spletne strani uspešno vključene v repozitorij portala Sistory. Vse statične spletne strani imajo tudi možnost dinamičnega prikazovanja vsebine.

Sustainability of digital editions

Use of static web pages at the History of Slovenia – Sistory portal

The contribution is based on the position that, with regard to digital editions, the highest possible degree of digital sustainability of data, presentations, functionalities, and programme code should be ensured. This represents a significant challenge, especially in case of smaller digital humanities projects with limited financing, which does not allow for the long-term maintenance of technically-demanding digital editions. The alternative solutions facilitated by the swift development of static web pages in the recent years are presented in the contribution. Digital editions based on the TEI have been successfully included in the Sistory portal repository as static web pages, employing basic XML (XSLT) and web technologies (HTML, CSS, JavaScript). All the static web pages also have the possibility of displaying dynamic content.

1. Uvod

V digitalni humanistiki je že dolgo časa prisotno zavedanje o pomembnosti digitalne trajnosti in trajnega ohranjanja digitalnih virov (Schaffner in Erway, 2014: 7). Raziskovalni podatki posameznega projekta ponavadi obstajajo dlje kot sam projekt, v okviru katerega so bili ti podatki zbrani, urejeni in objavljeni. Zato je zelo pomembno, da tudi po zaključku projekta poskrbimo za kvalitetno in trajnostno hrambo digitalnih podatkov.

V zadnjih letih se je intenzivno razpravljalo o tehničnih vidikih upravljanja z raziskovalnimi podatki in njihovega dolgoročnega arhiviranja: metapodatki, arhivski formati in mediji za hrambo, dokumentacija. Toda šele v zadnjem času smo se začeli zavedati, da je za kvalitetno upravljanje in ponovno uporabo teh podatkov skoraj še bolj pomembno ohranjanje podatkov v skladu s specifičnimi potrebami različnih znanstvenih disciplin. (Moeller et al., 2018) Medtem, ko v naravoslovnih in v družboslovnih vedah v glavnem uporabljajo podatke iz meritev in vprašalnikov, je v humanističnih vedah v prvem planu uporaba kulturnih objektov kot so rokopisi, besedila, slike in posnetki. V humanističnih raziskavah nato digitalne kulturne objekte pogosto še dodatno obdelajo, vizualizirajo, označijo, povežejo in interpretirajo (DHd-AG Datenzentrum, 2017: 7).

Takšen način obdelave podatkov je še zlasti pomemben pri digitalnih izdajah, ki so ključni del digitalne humanistike (Andorfer et al., 2016). Digitalne znanstvene izdaje so seveda v prvi vrsti predvsem raziskave, v okviru katerih nastajajo različni prepisi, označbe, analize, razlage ipd. Zato bi morali biti predvsem ti raziskovalni podatki dolgoročno in pod odprtimi pogoji dostopni raziskovalni skupnosti (Robinson, 2016). Tako je pri digitalnih izdajah kodirano besedilo najpomembnejši dolgoročni rezultat projekta. Zelo pomemben je tudi sam

prikaz, ki v okviru določene aplikacije predstavlja pogled projektne skupine na te podatke. Toda vsak tak pogled še zdaleč ni unikaten ali celo edini možen, temveč se lahko te podatke prikazuje na različne možne načine. (Turska et al., 2016) Z vsako novo interpretacijo se število možnih načinov prikazovanja še dodatno povečuje. Vsaka takšna prezentacija pa je nov raziskovalni rezultat, ki si prav tako zasluži dolgoročno hrambo.

Zato v humanističnih vedah rezultati raziskav niso samo raziskovalni podatki, temveč tudi predstavitev okolje in aplikacije, ki omogočajo interpretacijo podatkov, iskanje, filtriranje in brskanje po podatkih in njihovo povezovanje (DHd-AG Datenzentrum, 2017: 7). Če bi torej hranili samo raziskovalne podatke, bi bila prvotna prezentacija za vedno izgubljena, čeprav je tudi prezentacija sestavni del digitalne izdaje (Fechner, 2018). Obenem ne smemo pozabiti, da je tudi programska koda, ki smo jo uporabil za izdelavo digitalnih izdaj, prav tako kot digitalna izdaja tudi sestavni del znanstvene argumentacije (Andrews in Zundert, 2016).

Zato trajnostna hramba digitalnih izdaj predstavlja še toliko večji izziv. Pri tem se lahko digitalne izdaje glede na vsebino, izgled in funkcionalnosti med seboj zelo razlikujejo. Večinoma so rezultat specifičnega raziskovalnega projekta, ki ima na razpolago relativno omejene finančne in človeške vire. Ker člani projektne skupine, ki izhajajo iz humanističnih ved, po navadi nimajo ustreznih tehničnih znanj, se morajo pri tehničnem razvoju večinoma zanašati na zunanje izvajalce. Pri tem so digitalne izdaje odvisne od zelo hitrega razvoja spletnih tehnologij in standardov (Andorfer et al., 2016).

Ti izzivi glede trajnostne hrambe digitalnih izdaj se bodo s hitrim naraščanjem števila digitalnih izdaj v bodoče samo še dodatno okrepili (Fechner, 2018). Velik izziv je in bo predvsem za manjše digitalno humanistične projekte z omejenim financiranjem, ki ne omogočajo

dolgoročnega vzdrževanja tehnično zahtevnih digitalnih izdaj. Kot alternativo rešitev bom v nadaljevanju predstavil rešitve, ki jih ponuja hiter razvoj statičnih spletnih strani. V zadnjih nekaj letih so statične spletne strani postale eden glavnih trendov spletnega razvoja. Vse kaže, da se bo ta trend nadaljeval tudi v prihodnje.¹ V prispevku bom predstavil izkušnje, ki sem jih z generiranjem statičnih spletnih strani za digitalne izdaje pridobil v okviru dejavnosti Raziskovalne infrastrukture slovenskega zgodovinopisja, ki med drugim upravlja tudi portal Zgodovina Slovenije – Sistory.² Pri tem se bom omejil samo na statične spletne strani, ki jih generiramo iz XML datotek, kodirane v skladu s Smernicami Text Encoding Initiative (TEI) (TEI Consortium, 2018). Smernice TEI so namreč v digitalni humanistiki *de facto* standard za kodiranje besedil, ki ga uporablja veliko različnih humanističnih projektov in raziskav (Romary et al., 2017: 5).

V drugem poglavju bom najprej predstavil glavne prednosti in slabosti modernih statičnih spletnih strani. V našem primeru smo se odločili za nadgradnjo osnovnih pretvorb XSLT konzorcija TEI. V tretjem poglavju bom predstavil svojo generično nadgradnjo pretvorb XSLT konzorcija TEI in v četrtem poglavju projektno specifične možnosti njegove nadgradnje. V obeh teh poglavjih bom predstavil še različne možnosti dodajanja dinamične vsebine statičnim spletnim stranem. V zaključku bom omenil, kako te statične spletne strani vključujem v digitalno repozitorij portala Sistory in še naše načrte za bodoči razvoj.

2. Moderne statične spletne strani

Sprva so bile vse spletne strani statične, zaradi česar so bile seveda tudi vse digitalne izdaje s področja digitalne humanistike narejene kot statične HTML spletne strani. To je veljalo tudi za slovenske elektronske znanstvenokritične izdaje (Ogrin in Erjavec, 2009),³ ki so uvajale paradigmo digitalnih izdaj tudi v Sloveniji (Ogrin, 2005). Že kmalu so se ustvarjalci teh digitalnih izdaj srečali z nekaterimi pomanjkljivostmi takšnih statičnih spletnih strani. Predvsem so pogrešali možnost strukturiranega iskanja po besedilu, prilagodljivih parametrov prikaza in dinamičnega povezovanja vsebin. Zato so se pri novih digitalnih izdajah raje odločili uporabiti platformo Fedora Commons (Erjavec et al., 2011).

V tem času so na spletu že dolgo časa kraljevale dinamične spletne strani, ki so uspešno zamenjale zastarele statične spletne strani, pri katerih je bilo možno vsebino spreminjati le tako, da so razvijalci neposredno posegali v HTML kodo. Šele dinamične spletne strani so s pomočjo sistemov za upravljanje vsebin (npr. zelo popularni WordPress, Drupal in Joomla) naposled omogočile tudi tehnično nepodkovanim uporabnikom, da so lahko začeli na spletu objavljati zeleno vsebino.

Dinamične spletne strani imajo vsebino shranjeno v bazah podatkov. Strežnik vsebino zgradi šele takrat, ko odjemalec zahteva spletno stran in je kot takšna

prilagojena zahtevam uporabnika. Za komuniciranje s strežnikom se uporablja ustrezen programski jezik. Največji problem takšnih dinamičnih spletnih strani je ta, da so njene tehnične rešitve pogosto bolj zapletene od dejanskih potreb uporabnikov.

Moderne statične spletne strani so nastale kot odgovor na težave, ki pestijo dinamične spletne strani. V nasprotju z njimi statične spletne strani nimajo baz podatkov in strežniškega programskega jezika, temveč so zgolj skupek datotek HTML, CSS in JavaScript. Statične spletne strani imajo zato v primerjavi s dinamičnimi številne prednosti (Rinaldi, 2015):

- Zmogljivost: ker nimajo podatkovnih baz in procesiranja s strani strežnika, ni nevarnosti, da bi se takšne strani lahko upočasnile.
- Gostovanje: ker nimajo strežniškega programskega jezika, je gostovanje enostavno in poceni. Obstajajo celo zastojne možnosti kot so GitHub strani.
- Varnost: nimajo baz podatkov in strežniških programskih jezikov, ki bi jih nekdo lahko izkoristil za računalniške vdore. Zato so takšne strani varne, dokler so datoteke teh strani varno shranjene.
- Vzdrževanje: ker nimajo baz podatkov, strežniškega programskega jezika in sistema za upravljanje vsebin, je njihovo vzdrževanje zelo preprosto.
- Kontrola verzij: ker je celotna statična spletna stran sestavljena samo iz tekstovnih datotek, je vse njene verzije možno dokaj enostavno hraniti v sistemih za kontrolo verzij kot je npr. Git.

Ti razlogi so zlasti pomembni zaradi trajnosti digitalnih izdaj. Uporaba standardnih formatov kot so TIFF in JPEG za digitalne slike, HTML in XML za besedila ipd., ustvarjalcem digitalnih izdaj zagotavlja, da bodo njihove izdaje berljive in uporabne še dolgo časa (Turco, 2016). Zato so to paradigmo začeli poudarjati tudi v sorodnih projektih s področja digitalne humanistike (Viglianti, 2017; Daengeli in Zumsteg, 2017).

Ti razlogi pa so manj prepričljivi, če glede digitalnih izdaj pričakujemo, da bodo vsebovale tudi uporabniško generirano vsebino. Zato statične spletne strani niso primerne za vse digitalne izdaje s področja digitalne humanistike. V mnogih primerih njihove rešitve ne bodo mogle zadovoljiti potreb ustvarjalcev in uporabnikov. Po drugi strani pa je zelo veliko digitalnih projektov, kjer vsebina in njen prikaz nista tako zelo zahtevni. V teh primerih bi bile obstoječe rešitve, ki jih prinašajo statične spletne strani, več kot zadovoljive, predvsem zaradi tega, ker moderne statične strani niso povsem brez možnosti dodajanja dinamičnih vsebin. V resnici so statične spletne strani doživele svojo renesanso šele takrat, ko so se pojavile različne storitve in programske rešitve, ki so tem stranem omogočile dodajanje dinamičnih vsebin.

Modernih statičnih spletnih strani ne kodiramo več ročno, temveč jih generiramo s pomočjo generatorjev statičnih spletnih strani. Izbira teh generatorjev je danes zelo široka. Med najbolj uporabljenimi je Jekyll,⁴ ki se ga uporablja tudi pri izdelavi GitHub strani. Zato se je njegova uporaba razširila tudi na humanistiko (Visconti, 2016). Generatorji statičnih spletnih strani domnevajo, da bo uporabnik za pisanje vsebine uporabil besedilne

¹ Web Development Trends in 2018, <https://clockwise.software/blog/web-development-trends-in-2018/>.

² Raziskovalna infrastruktura slovenskega zgodovinopisja, <http://www.sistory.si/publikacije/?menuBottom=2>.

³ eZISS, <http://nl.ijs.si/e-zrc/>.

⁴ <https://jekyllrb.com/>.

formate kot je npr. med razvijalci zelo popularen Markdown.⁵ Te formate z generatorjem pretvorimo v HTML spletne strani in jih nato objavimo na spletu. Toda Markdown sintaksa je zelo pomanjkljiva in omogoča le osnovno objavljavanje vsebine, zaradi česar je neprimerna za označevanje kompleksnih humanističnih besedil. Zato besedila v humanistiki večinoma kodiramo za XML označevalnim jezikom. XSLT (Extensible Stylesheet Language for Transformation) pa uporabljamo kot orodje za pretvorbo XML. Skupaj tvorita ključni tehnologiji za digitalno humanistiko. (Flanders et al., 2016) Ker je uporaba pretvorb XSLT pogosto zelo podobna pretvorbam v generatorjih statičnih spletnih strani, lahko XSLT kljub njegovi dolgi tradiciji tudi označimo za »moderen, zmogljiv generator statičnih spletnih strani« (Kraetke in Imsieke, 2016).

3. Sistory TEI profil

Konzorcij TEI že dolga leta redno vzdržuje in posodablja pretvorne programe XSLT, s katerimi je iz dokumentov TEI mogoče generirati ne le (X)HTML spletne strani, temveč tudi številne ostale formate, mdr. LaTeX, XSL FO, ePub, DOCX in ODT. Ti pretvorni programi so odprto dostopni v GitHub repozitoriju in jih skupaj z novimi verzijami Smernic TEI redno posodablja z novimi verzijami.⁶ Nimajo samo dobre pisne dokumentacije,⁷ temveč je tudi programska koda vzorno komentirana. S pomočjo teh pretvornih programov tako med drugim generirajo tudi statične spletne strani vsakokratnih Smernic TEI.⁸

Predvsem pa pretvorbe XSLT konzorcija TEI preko profilov omogočajo zelo fleksibilno prilagoditev glede na potrebe posameznega projekta. Pretvorni programi TEI so bili v resnici napisani z namenom, da se jih lahko čim bolj prilagaja. Obstaja kopica parametrov, ki jih lahko konfiguriramo v skladu s svojimi željami. Stili vsebujejo številne XSLT spremenljivke in predloge, ki jih lahko predelamo v skladu s svojimi potrebami. Avtorji kode so celo pomislili na prazne (hook) predloge, katerim lahko dodamo svojo vsebino in programske kodo XSLT. Vse te možnosti sem izkoristil pri pisanju Sistory profila za pretvorbe XSLT konzorcija TEI.⁹

Sprva sem pri pisanju teh profilov izhajal iz potreb Raziskovalne infrastrukture slovenskega zgodovinopisja po fleksibilnem in čim bolj sprotnem objavljanju naše tehnične dokumentacije na spletu. V okviru raziskovalne infrastrukture s sodelavci med drugim upravljamo portal Zgodovina Slovenije – Sistory, ki vsebuje tudi repozitorij in digitalno knjižnico. Zato smo se odločili, da te digitalne izdaje po možnosti čim bolj intenzivno vključimo v obstoječo infrastrukturo. Do leta 2016 smo tako statične spletne strani digitalnih izdaj hranili na dodatnem www2 strežniku portala Sistory,¹⁰ v sami digitalni knjižnici smo hranili samo metapodatke o digitalnih izdajah in povezave na te statične spletne strani. Z nadgradnjo portala Sistory v letu 2016 smo HTML in vse druge datoteke teh

digitalnih izdaj lahko začeli hraniti neposredno v repozitoriju in digitalni knjižnici.

Zaradi želje po čim večji integraciji digitalnih izdaj v portal Sistory sem tudi zunanji izgled digitalnih izdaj poskusil čim bolj približati uporabniškemu vmesniku portala. Na sliki 1 je kot primer prikazan posnetek vstopne strani portala med leti 2012 in 2016 ter na sliki 2 uporabniški vmesnik digitalne izdaje iz leta 2014.



Slika 1: Vhodna spletna stran portala Sistory iz leta 2016



Slika 2: Uporabniški vmesnik digitalne izdaje iz leta 2014

Čprav so barvni toni povsem isti in se tudi postavitve logotipa, iskalne vrstice, glavne zgornje navigacije in vsebine zelo dobro zgledujejo po portalu Sistory, uporabniška vmesnika le nista povsem ista. Uporabniški vmesnik takratnega portala je bil namreč še vedno zgrajen na osnovi stare HTML 4 tehnologije, medtem ko sem pri digitalnih izdajah že začel uporabljati odziven dizajn spletnih strani in HTML 5. Pri tem sem se odločil za uporabo ogrodja za odzivne uporabniške vmesnike ZURB Foundation.¹¹ Moje prilagoditve in dodatke CSS in JS hranim v GitHub repozitoriju.¹² Ker se je uporaba tega ogrodja izkazala za zelo koristno, smo ga leta 2016 vključili tudi v prenovljen portal Sistory. Novemu izgledu portala sem nato prilagodil še izgled digitalnih izdaj (primerjaj sliki 3 in 4).

⁵ <http://www.daringfireball.net/projects/markdown/>.

⁶ TEI XSL Stylesheets, <https://github.com/TEIC/Stylesheets>.

⁷ XSL stylesheets for TEI XML, <http://www.tei-c.org/release/doc/tei-xsl/>.

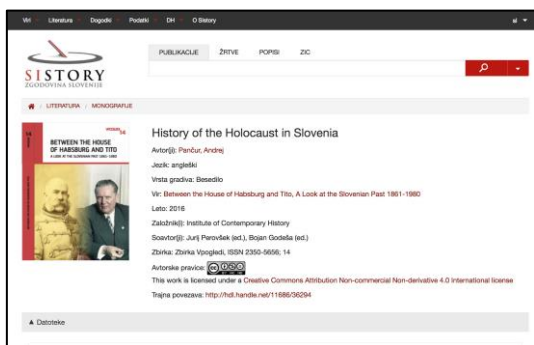
⁸ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

⁹ Sistory TEI XSL Stylesheets, <https://github.com/Sistory/Stylesheets>.

¹⁰ <http://www2.sistory.si/>.

¹¹ Foundation, <https://foundation.zurb.com/>.

¹² Sistory themes, <https://github.com/Sistory/themes>.



Slika 3: Zgornja navigacija, iskalna vrstica in metapodatkovna stran portala Sistory



Slika 4: Uporabniški vmesnik digitalne izdaje iz leta 2016

Poleg prvotno načrtovane tehnične dokumentacije smo na spletu v formatu HTML kmalu začeli objavljati še druge vrste publikacij, predvsem monografije, zbornike in revije. Zato sem Sistory TEI profil konfiguriral v skladu s potrebami po objavljanju takšnih vrst digitalnih izdaj. Profil omogoča pretvorbo:

- posameznega TEI dokumenta;
- več TEI dokumentov iz skupnega TEI korpusa. V tem primeru je potrebno vsak TEI dokument pretvoriti posebej. Posebej je potrebno pretvoriti še sam TEI korpus in njegov <teiHeader>, saj na ta način generiramo skupno naslovnico, kolofon in kazala vsebine.

Glavna menijska navigacija po digitalni izdaji se nahaja povsem na vrhu spletne strani kot horizontalna navigacija s spustnim menijem. Struktura te navigacije odraža strukturo ter sklope in razdelke posameznih dokumentov TEI. V nadaljevanju bom na kratko predstavil možne vsebinske sklope tako navigacije kot dokumenta TEI. V praksi seveda noben dokument nima prav vseh teh razdelkov, temveč si jih avtor dokumentov TEI oblikuje povsem v skladu s svojimi potrebami.

Osrednji del vsebine je vedno znotraj <body> elementa. Glavna vsebina mora biti nujno v enem ali več <div> elementih, ki morajo obvezno imeti atribut @xml:id. Vsak tak <div> predstavlja svoj razdelek vsebine oziroma poglavja. Zato so v navigaciji vsi <div> znotraj <body> prikazani v enem spustnem meniju. Pred in za tem spustnim menijem je lahko dostopna še različna ostala vsebina, ki je v TEI dokumentu kodirana znotraj <front> in <back> elementov. V sliki 5 so tako prikazani vsi ti glavni vsebinski sklopi.

```
<text>
  <front>
    <titlePage>
      <docTitle>
        <titlePart>Naslov</titlePart>
      </docTitle>
      <docAuthor>Avtor</docAuthor>
      <docEdition>Izdaja</docEdition>
      <docImprint>
        <pubPlace>Kraj izdaje</pubPlace>
        <docDate>Datum izdaje</docDate>
      </docImprint>
      <graphic url="url_do_slike.jpg"/>
    </titlePage>
    <div type="preface" xml:id="prf-01">
      <!-- Uvodna poglavja -->
    </div>
  </front>
  <body>
    <div type="chapter" xml:id="ch01">
      <!-- Poglavja z glavno vsebino -->
    </div>
  </body>
  <back>
    <div type="bibliogr" xml:id="bibl01">
      <!-- Bibliografije -->
    </div>
    <div type="appendix" xml:id="app01">
      <!-- Priloge -->
    </div>
    <div type="summary" xml:id="sum01">
      <!-- Povzetki -->
    </div>
  </back>
</text>
```

Slika 5: Glavni vsebinski sklopi dokumenta TEI

Med njimi je obvezen le <titlePage>, ki je izhodiščni index.html element in je kot takšen v navigaciji dostopen na prvem mestu kot Naslovnica. Znotraj <front> se lahko nahajajo eden ali več <div> elementov, ki v navigaciji predstavljajo sklop uvodnih poglavij. Znotraj <back> elementa imamo tri možne vsebinske sklope (bibliografije, priloge, povzetki), zaradi česar je nujno, da imajo vedno ustrezen atribut @type. Vsak ta sklop ima lahko enega ali več poglavij. V večini primerov poteka pretvorba vsebine teh razdelkov na podlagi standardnih pretvorb XSLT konzorcija TEI, ki sem jih sem za potrebe naših digitalnih izdaj le delno prilagodil. Povsem na novo sem napisal pretvorbe za generirane razdelke <divGen>. Vsi so vključeni v Sistory TEI profil. Ti generirani razdelki so lahko vključeni v <front> (Slika 6) ali <back> (Slika 7). Vsak <divGen> mora vsebovati <head> s poljubnim naslovom razdelka. Ti naslovi so nato vključeni v navigacijo spletne izdaje.

Za razliko od zgoraj omenjenimi razdelkov <div>, kjer so identifikatorji @xml:id samo priporočljivi (HTML datoteke teh razdelkov dobijo imena po teh identifikatorjih), so pri generiranih razdelkih nujni in imajo tudi semantičen pomen, ki je ključen za njihovo pretvorbo. Atribut @type opredeljuje glavno kategorijo, ki je v horizontalni navigaciji posebej izpostavljena. Atribut @xml:id bolj natančno opredeljuje podkategorijo, ki je v navigaciji prikazana v spustnem meniju. Najobsežnejša kategorija je skupina kazal vsebine (Table of Contents TOC), ki poleg različnih kazal vsebine poglavij in podpoglavij vsebuje še kazalo tabel, slik in grafikonov. Kazalo grafikonov je v resnici posebna skupina kazal slik

(<figure>), ki zajema slike z atributom @type in vrednostjo chart.

```
<front>
  <divGen type="cip" xml:id="cip">
    <head>Kolofon CIP</head>
  </divGen>
  <divGen type="teiHeader" xml:id="teiHeader">
    <head>TEI kolofon</head>
  </divGen>
  <divGen type="toc" xml:id="id-toc">
    <head>Kazalo vsebine</head>
  </divGen>
  <divGen type="toc" xml:id="id-images">
    <head>Kazalo slik</head>
  </divGen>
  <divGen type="toc" xml:id="id-charts">
    <head>Kazalo grafikonov</head>
  </divGen>
  <divGen type="toc" xml:id="id-tables">
    <head>Kazalo tabel</head>
  </divGen>
  <divGen type="toc" xml:id="id-titleAuthor">
    <head>Kazalo vsebine, kjer izpiše še
      ime avtorja</head>
  </divGen>
  <divGen type="toc" xml:id="id-titleType">
    <head>Kazalo vsebine</head>
  </divGen>
  <divGen type="search" xml:id="search">
    <head>Iskalnik</head>
  </divGen>
</front>
```

Slika 6: Seznam vseh možnih generiranih razdelkov
<divGen> v <front>

V okviru elementa <back> se nahaja samo ena kategorija generiranih razdelkov, ki zajema različne sezname oziroma indekse oseb, krajev in organizacij. Generirani razdelki zajemajo vse osebe iz dokumenta TEI kodirane z elementov <persName> ali vse kraje <placeName> ali vse organizacije <orgName>. Vse tako kodirane imenske entitete morajo imeti atribut @ref, preko katerega se sklicujejo na ustrezni kanonični element v seznamu entitet (<listPerson> za osebe, <listOrg> za organizacije in <listPlace> za kraje) v glavi dokumenta TEI (<teiHeader>). Atribut @ref elementa <placeName> lahko vsebuje tudi sklic na GeoNames¹³ ali DBpedia¹⁴ URI, kjer Sistory profil poišče geografske koordinate, ki jih nato prikaže v seznamu krajev.

Ker je s Sistory profilom mogoče pretvarjati tudi dokumente TEI iz TEI korpusa, <divGen> iz različnih dokumentov TEI ne morejo imeti istih @xml:id identifikatorjev. Zato se podkategorije generiranih razdelkov določi tako, da je identifikator podkategorije zapisan za zadnjim pomišljajem vrednosti tega identifikatorja (glej sliko 6 in 7, kjer vrednost id pred pomišljajem v @xml:id opredeljuje poljubni identifikator, za pomišljajem pa podkategorijo).

Sistory profil omogoča tudi prikaz dinamične vsebine. Kot osnovna funkcionalnost je vključen iskalnik Tipue Search.¹⁵ Vključimo ga z generiranim razdelkom tipa

search v <front>. Tipue Search je odprtokodni jQuery vtičnik, ki ga je mogoče relativno enostavno vključiti tudi v statične spletne strani. Na grafičnem vmesniku je iskalna vrstica postavljena takoj pod spodnjo navigacijo, <divGen> iskalnika pa generira search.html spletno stran, ki vključuje dinamičen prikaz rezultatov iskanja. Vsebina dokumenta TEI je kot JavaScript objekt (JSON) indeksirana v datoteki tipuesearch_content.js, ki mora biti v istem direktoriju kot datoteka search.html. Indeksacija vsebine poteka na ravni odstavkov <p>, seznamov <list>, tabel <table>, slik <figure> in vseh drugih možnih elementov TEI, ki so neposredni child elementi razdelka <div>. Zato morajo vsi ti elementi imeti identifikator @xml:id. Edina izjema so sezname. Če le-ti nimajo atributa @xml:id, imajo pa jih njihovi child elementi, potem so indeksirani slednji.

```
<back>
  <divGen type="index" xml:id="id-persons">
    <head>Seznam oseb</head>
  </divGen>
  <divGen type="index" xml:id="id-places">
    <head>Seznam krajev</head>
  </divGen>
  <divGen type="index" xml:id="id-organizations">
    <head>Seznam organizacij</head>
  </divGen>
</back>
```

Slika 7: Seznam vseh možnih generiranih razdelkov
<divGen> v <back>

4. Konfiguracija in nadgradnja Sistory profila

Tako kot glavni pretvorni programi XSLT konzorcija TEI je tudi Sistory profil narejen z namenom, da ga je mogoče prilagoditi potrebam posameznega projekta. V ta namen vključuje nekaj izvornih parametrov pretvorb XSLT konzorcija TEI, ki sem jim dodal še nekaj novih parametrov. Vse te parametre je sicer mogoče na novo nastavljeni ob vsaki pretvorbi, vendar je bolj priporočljivo, da za vsak projekt naredimo nov projektni profil. Običajno pretvorba poteka tako, da projektni XSLT profil vključi Sistory XSLT profil, ta pa vključi TEI pretvorbe XSLT.

Privzeti Sistory profil tako npr. predvideva, da bo pri pretvorbi vsako poglavje oziroma prvi razdelek <div> samostojna HTML spletna stran. V tem primeru se na spletnih straneh avtomatično doda navigacija z gumbi naprej in nazaj. Za razliko od prvotnih pretvorb TEI ta navigacija vključuje tudi generirane razdelke <divGen>. Toda s spremembo parametra splitLevel (izvorno parameter pretvorb TEI) lahko določimo, da so tudi podpoglavja ločene HTML spletne strani. Temu primerno je sedaj prilagojena tudi navigacija naprej/nazaj, navzgor/navzdol med spletnimi stranmi. Trenutno Sistory profil podpira le globino treh razdelkov.

S parametrom documentationLanguage je trenutno mogoče nastaviti slovensko, angleško in srbsko navigacijo (latinica ali cirilica). Z dodajanjem novih prevodov v dokument myi18n.xml je mogoče to lokalizacijo še razširiti. Ustrezno je poskrbljeno tudi za lokalizacijo iskalnika Tipue Search.

Sistory profil omogoča še vzporedno objavo različnih jezikovnih verzij besedila. V tem primeru morajo imeti vsi glavni razdelki <div> in generirani razdelki <divGen> atributa @xml:lang z ustrezno jezikovno kodo in @corresp s kazalko na vse drugojezične verzije tega

¹³ GeoNames, <http://www.geonames.org/>.

¹⁴ DBpedia, <http://wiki.dbpedia.org/>.

¹⁵ Tipue Search, <http://www.tipue.com/search/>.

besedila. Obenem je potrebno še nastaviti parameter `languages-locale` na vrednost `true` in s parametrom `languages-locale-primary` določiti jezikovno kodo izhodiščne datoteke `index.html`.

Podobno prilagodljiv je tudi prikaz vseh metapodatkov dokumenta TEI iz `<teiHeader>`. To pretvorbo najprej določimo s vključitvijo generičnega razdelka tipa `teiHeader` (glej sliko 6). Celotna vsebina `<teiHeader>` je pretvorjena v HTML elemente seznam definicij `<dl>`, pri katerem oznaka definicije `<dt>` opredeljuje ime TEI elementa ter imena in vrednosti atributov (element `[atribut = vrednost | atribut = vrednost]`), definicija elementa `<dd>` pa vsebino besedila elementa. Definicije so seveda temu primerno tudi gnezdene. S dodanimi parametri lahko to pretvorbo konfiguriramo tako, da namesto imen elementov in atributov izpiše njihova opisna imena v angleščini ali slovenščini.

Poleg te preproste konfiguracije Sistory profila lahko pri projektni pretvorbi vključimo še kakršno koli dodatno pretvorbo XSLT, ki jo tako povsem prilagodimo potrebam digitalne izdaje. Obenem lahko z vključitvijo različnih JavaScript knjižnic in vtičnikov ter spletnih aplikacij omogočimo še dodatno dinamično prikazovanje vsebine. V primeru digitalnih izdaj na portalu Sistory sem npr. za prikazovanje večje količine tabelarnih podatkov uspešno uporabil `DataTables`,¹⁶ za grafikone `Highcharts`,¹⁷ za zemljevide `Google Maps`, za slike `ImageViewer`.¹⁸ In to so samo nekateri primeri, ki imajo tudi različne alternative. Vsako leto pa se jim pridružijo še številne nove možnosti.

Obenem so se leta 2017 z objavo `Saxon-JS`¹⁹ še dodatno izboljšale možnosti dinamičnega prikazovanja vsebine XML dokumentov v statičnih spletnih straneh. To možnost sem že uspešno uporabil pri spletnih seznamih, kjer sem prikaz celotne vsebine seznama filtriral glede na želeni parameter (identifikator).

5. Zaključek

Privzeta pretvorba Sistory profila generira vse HTML, JS in morebitne ostale datoteke v isti direktorij. Ker je tako generirana digitalna izdaja sestavljena zgolj iz statičnih spletnih strani, jo lahko uporabljamo tudi na osebнем računalniku. Na ta način je možno digitalno izdajo tudi učinkovito testirati še pred objavo na spletu, kjer jo lahko hitro in enostavno objavimo na poljubnem dostopnem strežniku. Kot zastojna možnost obstajajo tudi spletne strani `GitHub` repozitorija. Na ta način zagotovimo še učinkovito kontrolo verzij.

Toda glavni namen Sistory profilov je vključitev digitalnih izdaj neposredno v repozitorij portala Sistory in njegovo digitalno knjižnico. Na ta način lahko zagotovimo ne le učinkovito hrambo vseh datotek digitalne izdaje, skupaj z dodajanjem unikatnih `handle` identifikatorjev in kontrolne vsote (`checksum`) za vse datoteke, temveč tudi fleksibilno razvrščanje digitalne izdaje v enega ali več digitalnih objektov, z enim ali več intelektualnih entitet. Vsaka intelektualna entiteta ima svoj `Handle` identifikator in svoje metapodatke. Vključuje lahko nič ali več datotek. Datoteke posamezne intelektualne entitete se nahajajo v istem direktoriju. Pot do direktorija vsebuje tudi predpono

(suffix) `Handle` identifikatorja, ki je v primeru portala Sistory vedno številčna vrednost (npr. za predpono 555 je relativna pot `/cdn/publikacije/1-1000/555/datoteka`). Sistory XSLT profil mora zato nujno že vnaprej natančno vedeti, kakšne so vrednosti teh identifikatorjev. Tako lahko tudi vnaprej natančno določimo, ali bo celotna vsebina digitalne izdaje samo v eni intelektualni entiteti portala Sistory ali pa bodo različne datoteke digitalne izdaje vključene v različne intelektualne entitete. Te identifikatorje zapišemo med ostale metapodatke v `<teiHeader>` in sicer v okviru `<publicationStmt>` kot vrednost enega ali več elementov `<idno>`. Ta element mora imeti vrednost atributa `@type` `sistory`, atribut `@corresp` pa mora imeti kazalke na vse ustrezne razdelke `<div>` in `<divGen>`, katerih vsebina bo vključena v intelektualno entiteto s tem identifikatorjem.

Sistory XSLT profil je pod odprtimi pogoji dostopen v `GitHub` repozitoriju.²⁰ V drugem `GitHub` repozitoriju so dostopne še vse digitalne izdaje, ki se nahajajo na portalu Sistory. Za vsako teh izdaj so dostopne še projektna nadgradnje Sistory XSLT profila.²¹ Profil redno dopolnjujem in seveda tudi vzdržujem v skladu s spremembami pretvornih programov XSLT konzorcija TEI. V letošnjem letu načrtujem še obsežnejšo nadgradnjo. Trenutni portal Sistory bomo namreč v naslednjem letu zamenjali z novim repozitorijem, ki ga pravkar intenzivno razvijamo. Novi repozitorij bo v okviru METS aplikacijskega profila omogočil fleksibilno dodajanje različnih vrst metapodatkov, med drugim tudi `teiHeader`. S tem bomo občutno izboljšali metapodatkovno opremljenost digitalnih izdaj na portalu Sistory.

6. Zahvala

Raziskavo je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije skozi program Raziskovalne infrastrukture slovenskega zgodovinopisja (I0-0013) in slovenske raziskovalne infrastrukture DARIAH-SI.

7. Literatura

Peter Andorfer, Matej Āurčo, Thomas Stacker, Christian Thomas, Vera Hildenbrandt, Hubert Stigler, Sibylle Soring in Lukas Rosenthaler. 2016. Nachhaltigkeit technischer Losungen fur digitale Editionen: Eine kritische Evaluation bestehender Frameworks und Workflows von und fur Praktiker_innen. V: *DHD 2016: Modellierung – Vernetzung – Visualisierung: Die Digital Humanities als facherübergreifendes Forschungsparadigma: Konferenzabstracts*, str. 36-39. Universitat Leipzig. <http://www.dhd2016.de/>.

Tara Andrews in Joris van Zundert. 2016. What Are You Trying to Say? The Interface as an Integral Element of Argument. V: *Digital Scholarly Editions as Interfaces*, International Symposium at the University of Graz, Austria, str. 31-32. Centre for Information Modelling – Austrian Centre for Digital Humanities. https://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/dse-interfaces_BoA21092016.pdf.

¹⁶ `DataTables`: Table plug-in for `jQuery`, <https://datatables.net/>.

¹⁷ `Highcharts`, <https://www.highcharts.com/products/highcharts/>.

¹⁸ `ImageViewer`, <http://ignitersworld.com/lab/imageViewer.html>.

¹⁹ `Saxon-JS`, <http://www.saxonica.com/saxon-js/index.xml>.

²⁰ Sistory TEI XSL Stylesheets,

<https://github.com/Sistory/Stylesheets>.

²¹ <https://github.com/Sistory/publications>.

- Peter Daengeli in Simon Zumsteg. 2017. Hermann Burgers Lokalbericht: Hybrid-Edition mit digitalem Schwerpunkt. V: *DHd 2017: Digitale Nachhaltigkeit: Konferenzabstracts*. Universität Bern. str. 151-155. <http://www.dhd2017.ch/>.
- DHd-AG Datenzentren. 2017. *Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum*, Grundsatzpapier zur Sicherung der langfristigen Verfügbarkeit von Forschungsdaten. Hamburg. DOI: 10.5281/zenodo.1134760
- Tomaž Erjavec, Jan Jona Javoršek, Matija Ogrin in Petra Vide Ogrin. 2011. Od biografskega leksikona do znanstvenokritične izdaje: vprašanje trajnosti elektronskih besedil. *Knjižnica*, 55(1): 103-114. <https://knjiznica.zbds-zveza.si/knjiznica/article/view/6004>.
- Martin Fechner. 2018. Eine nachhaltige Präsentationsschicht für digitale Editionen. V: Georg Vogeler (ur.), *DHd 2018: Kritik der digitalen Vernunft: Konferenzabstracts*, str. 203-207. Universität zu Köln. <http://dhd2018.uni-koeln.de/>.
- Julia Flanders, Syd Bauman in Sarah Connell. 2016. XSLT: Transforming our XML data. V: C. Crompton, R. J. Lane in R. Siemens (ur.), *Doing Digital Humanities: Practice, Training, Research*, str. 255-272. Oxon in New York, Routledge.
- Martin Kraetke in Gerrit Imsieke. 2016. XSLT as a Modern, Powerful Static Website Generator: Publishing Hogrefe's Clinical Handbook of Psychotropic Drugs as a Web App. V: *Proceedings of XML in, Web Out: International Symposium on sub rosa XML*, Balisage Series on Markup Technologies, vol. 18. <https://doi.org/10.4242/BalisageVol18.Kraetke02>.
- Katrin Moeller, Matej Đurčo, Barbara Ebert, Marina Lemaire, Lukas Rosenthaler, Patrick Sahle, Ulrike Wuttke in Jörg Wettlaufer. 2018. Die Summe geisteswissenschaftlicher Methoden? Fachspezifisches Datenmanagement als Voraussetzung zukunftsorientierten Forschens. V: Georg Vogeler (ur.), *DHd 2018: Kritik der digitalen Vernunft: Konferenzabstracts*, str. 89-93. Universität zu Köln. <http://dhd2018.uni-koeln.de/>.
- Matija Ogrin in Tomaž Erjavec. 2009. Ekdotika in tehnologija: Elektronske znanstvenokritične izdaje slovenskega slovstva. *Jezik in slovstvo*, 54(6): 57-72. <http://www.dlib.si/?URN=URN:NBN:SI:doc-BOC8BANS>.
- Matija Ogrin (ur.). 2005. *Znanstvene razprave in elektronski mediji: razprave*. Ljubljana: Založba ZRC, ZRC SAZU. <http://nl.ijs.si/e-zrc/bib/eziss-knjiga.pdf>.
- Brian Rinaldi. 2015. *Static Site Generators: Modern Tools for Static Website Development*. Sebastopol, CA: O'Reilly Media.
- Peter Robinson. 2016. Why Interfaces Do Not and Should Not Matter for Scholarly Digital Editions. V: *Digital Scholarly Editions as Interfaces*, International Symposium at the University of Graz, Austria, str. 29-30. Centre for Information Modelling – Austrian Centre for Digital Humanities. https://static.uni-graz.at/fileadmin/gewi-zentren/Informationsmodellierung/PDF/dse-interfaces_BoA21092016.pdf.
- Laurent Romary, Piotr Banski, Jack Bowers, Emiliano Degl'innocenti, Matej Đurčo, Roberta Giacomi, Klaus Illmayer, Adeline Joffres, Fahad Khan, Mohamed Khemakhem, et al. 2017. *Report on Standardization (draft)*. [Technical report] 4.2 Inria. <https://hal.inria.fr/hal-01560563>.
- Jennifer Schaffner in Ricky Erway. 2014. *Does Every Research Library Need a Digital Humanities Center?* Dublin, Ohio: OCLC Research. <https://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-digital-humanities-center-2014.pdf>.
- TEI Consortium, ur. 2018 TEI P5: *Guidelines for Electronic Text Encoding and Interchange 3.3.0*. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>.
- Roberto Rosselli Del Turco. 2016. The Battle We Forgot to Fight: Should We Make a Case for Digital Editions? V: Matthew James Driscoll in Elena Pierazzo (ur.), *Digital Scholarly Editing: Theories and Practices*, str. 219-238. Cambridge, UK: Open Book Publishers. <http://dx.doi.org/10.11647/OBP.0095>.
- Magdalena Turska, James Cummings in Sebastian Rahtz. 2016. Challenging the Myth of Presentation in Digital Editions. *Journal of the Text Encoding Initiative*, (9). DOI: 10.4000/jtei.1453
- Raffaele Vigiante. 2017. Your own Shelley-Godwin Archive: An off-line strategy for an online publication (poster). V: *TEI 2017 Victoria*. https://hcmc.uvic.ca/tei2017/abstracts/t_126_vigliante_s_helleygodwin.html.
- Amanda Visconti. 2016. Building a static website with Jekyll and GitHub Pages. *The Programming Historian*, 5. <https://programminghistorian.org/lessons/building-static-sites-with-jekyll-github-pages>.