

Tamara Butigan-Vučaj
(Narodna biblioteka Srbije, Beograd)

Dejan Jovanović, student
(Matematički fakultet, Beograd)

TESTIRANJE SOFTVERA ZA DIGITALNE KOLEKCIJE ***GREENSTONE***

Narodna biblioteka Srbije i Nacionalni centar za digitalizaciju prvi put su saradivali na jednom konkretnom poslu, na testiranju poznatog softvera otvorenog koda za digitalne biblioteke pod nazivom GREENSTONE. Napravljeno je nekoliko digitalnih kolekcija u ovom softveru i mogu se pogledati na adresi <http://toledo.mi.sanu.ac.yu/gsd/cgi-bin/library>.

Ključne reči: *Greenstone*, digitalne kolekcije, digitalni dokumenti

Narodna biblioteka Srbije planira razvoj svoje digitalne forme kroz nastojanje da konstituiše paralelnu virtuelnu instituciju pod imenom *Digitalna Narodna biblioteka Srbije*. Preduslov za takvu formu jeste postojanje digitalnih kolekcija, koje su sinonim za digitalnu biblioteku u užem smislu. Fizički posao oko digitalne konverzije izabrane građe uveliko je u toku, još uvek se razmišlja o formatima digitalnih dokumenata, načinu formiranja digitalnih kolekcija i pristupu digitalnom materijalu.

Nacionalni centar za digitalizaciju i Narodna biblioteka Srbije po prvi put su se našli na zajedničkom poslu testiranja novozelandskog softvera otvorenog koda za digitalne biblioteke pod nazivom *Greenstone*. Ovaj softver je primenjen u mnogim digitalnim bibliotekama širom sveta, potpuno je besplatan (instalacija se može preuzeti sa zvaničnog web sajta ovog projekta <http://www.greenstone.org/>) i podržava najrazličitije formate dokumenata. Ideja je bila da se softver testira za potrebe Narodne biblioteke Srbije s perspektivom da se pokrene nacionalni projekat za pisanje softvera za digitalne biblioteke.

Softver *Greenstone* razvija se na Vaikatoskom Univerzitetu u Hamiltonu na Novom Zelandu, a deo je projekta Digitalne biblioteke Novog Zelanda, koji je podržao UNESCO. Digitalne zbirke ove biblioteke kao i osnovne informacije o projektu mogu se naći na lokaciji <http://www.nzdl.org/>. Prva namena ovog projekta bila je da pruži humanitarne i slične informacije zemljama u razvoju. Digitalne kolekcije se ovim softverom jednostavno mogu izvoziti na CD, što je naročito praktično ako nema Interneta ili su veze loše. U istom formatu, kolekcije su dostupne na internetu. Ovaj softver ima mogućnost ugradnje servera, i ako se sistem poveže na Intranet, informacije su dostupne svim računarima u mreži, što je posebno pogodno za škole, bolnice, biblioteke...

Kreatori ovog softvera definisali su osnovne kategorije vezane za digitalne biblioteke. Digitalna biblioteka je fokusirana kolekcija digitalnih objekata (tekstualnih, video i audio) zajedno s metodama za pristup i pronalaženje, kao i za izbor, organizaciju i održavanje digitalnih objekata. Digitalna zbirka podrazumeva nekoliko hiljada (ili miliona) dokumenata. Digitalni dokument je bilo koji skup informacija zabeležen u elektronskom formatu: tekst, slika, zvučni ili video zapis. “Bildovanje” ili kreiranje kolekcije je proces formiranja strukture za pristup informacijama u kolekciji. “Rebildovanje” kolekcije je proces kojim se može uvrstiti novi materijal u već postojeću kolekciju.

Softver *Greenstone* omogućava pristup kolekcijama preko pretraživača (lokalni i udaljeni). Radi pod operativnim sistemomima Windows i Unix. Omogućava unos različitih formata dokumenata pomoću «*plug-in*»-ova za: HTML, DOC, PDF, RTF, EML, PostScript, PowerPoint, Excel dokumente. Pretraživanje je moguće po punom tekstu i po indeksima: autora, naslova, poglavlja. Pregledanje sadržaja zbirki rešeno je preko listi: autora, naslova, datuma kreiranja, itd. Softver podržava *Unicode*, te je moguće upravljati dokumentima na ma kom jeziku, a takođe i birati jezik korisničkog interfejsa. Može da podrži tekst, slike, audio i video klipove. Projektovan je za multigigabajtne sisteme. Ugrađena je mogućnost izvoženja zbirki na CD. Ovaj softver podržava takozvane «distribuirane kolekcije»: protokol Z39.50 u potpunosti je podržan i za pristup eksternim serverima i za predstavljanje kolekcija eksternim korisnicima, a sada i Inicijativa otvorenih arhiva.¹

Digitalni dokumenti mogu biti prezentovani na više načina. Mogu biti hijerarhijski strukturirani, što zahteva dodatnu pripremu ulaznog HTML dokumenta. Mogu se prikazati u punom HTML tekstu, kao nestrukturirani. Mogu da se postave skenirane slike teksta, što je memorijski zahtevnije, ali u nekim situacijama sasvim pogodno. Uz skenirane slike može ići i prepoznati tekst, kako bi se imao izvorni izgled dokumenta, paralelno uz pretraživi tekst. Kolekcije se mogu kreirati od audio dokumenata i fotografija, kao i od video dokumenata. Što se tiče štampanih muzikalija, one se mogu postaviti kao skenirane slike i kao pretraživi dokumenti, pošto se obrade nekim od programa za optičko prepoznavanje muzike (OMR). Dokumenti mogu biti na različitim jezicima i pismima.

Greenstone ima mogućnost pridruživanja metapodataka uz digitalne dokumente. Podržani su različiti standardi za metapodatke: MARC, Dublin Core, BibTeX, Refer (za tehničke tekstove), kao i metapodaci za slike i multimedije kao TIFF i MPEG-7. Pretraživanje nudi dva tipa upita: Bulovo pretraživanje i rangiranje, a opcije za upit se biraju između jednostavnog, naprednog i pretraživanja s istorijom. Već je pomenuta mogućnost pregledanja lista dokumenata. One mogu biti formirane alfabetski prema naslovima, prema datumu kreiranja, po nekom od klasifikacionih sistema (UDK).

Softver *Greenstone* je široko primenjen i važi za jedan od najpoznatijih softvera za digitalne biblioteke. UNESCO organizuje obuke u zemljama Afrike i Azije. Nekoliko primera primene:

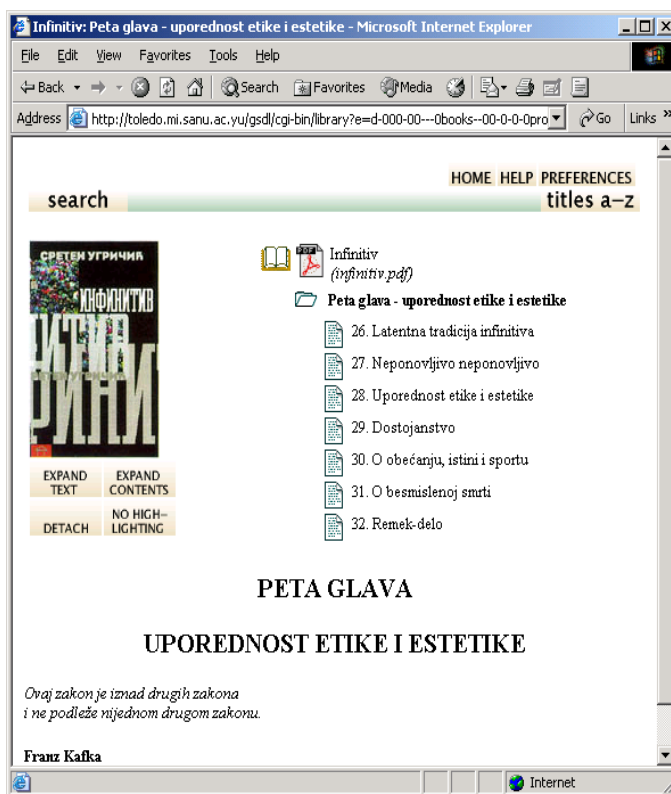
- Univerzitet UN u Tokiju <http://www.unu.edu/unupress/cdrom.html>
- Projekat Gutenberg <http://promo.net/pg/>
- Digitalna biblioteka Novog Zelanda <http://www.nzdl.org>
- Šopenovi rani radovi <http://chopin.lib.uchicago.edu/>
- Digitalna biblioteka na ruskom jeziku <http://gov.mari.ru/gSDL/cgi/library>

¹ Open Archive Initiative (OAI)

- Test Narodne biblioteke Srbije <http://toledo.mi.sanu.ac.yu/gsd/cgi-bin/library>

Testna digitalna biblioteka Narodne biblioteke Srbije sadrži tri zbirke: to su *Knjige*, časopis *Zenit* i *Rukopisi*. Interfejs je delom preveden na srpski, jezik interfejsa bira se u podešavanjima, pošto se uđe u određenu zbirku.

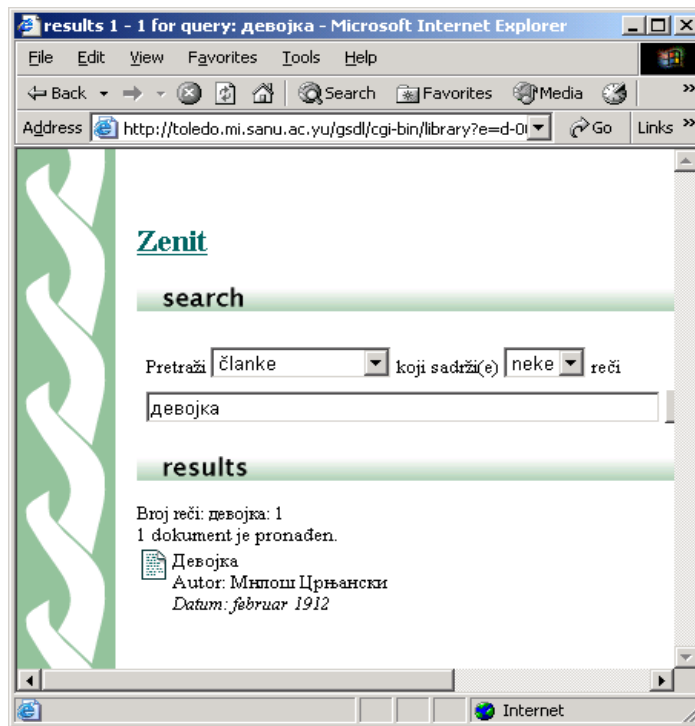
Zbirka *Knjige* sadrži dve publikacije, a obe su dostupne u formatima HTML i PDF. Puni tekst (HTML) je strukturiran i daje linkove na poglavlja knjige, te se knjiga može čitati po delovima. Ispod naslova ili slike korica, su ponuđene mogućnosti za rukovanje e-tekstom. Može se prikazati čitav tekst knjige na ekranu, može se otvoriti tekst u novom prozoru pretraživača, postoji i mogućnost da rezultati pretraživanja budu osvetljeni u tekstu. Format PDF dodat je *plug-in*-om, da bi se imao originalni izgled knjiga i da bi se one mogle preuzeti s mreže. Pretraživanje je po ključnim rečima jer nisu ugrađeni metapodaci za ove dokumente. Može se pretraživati po celim knjigama, po sekcijama ili po naslovima sekcija. Pretražuje se cela zbirka odjednom. Knjigama se najjednostavnije može pristupiti preko abecedne liste naslova. Na slici 1 dat je prikaz ekrana s poglavljima pete glave knjige *Infinitiv*. Otvoren je samo uvodni tekst, a tekstovi pojedinih poglavlja moraju se najpre otvoriti.



Slika 1. Prikaz strukturiranog teksta

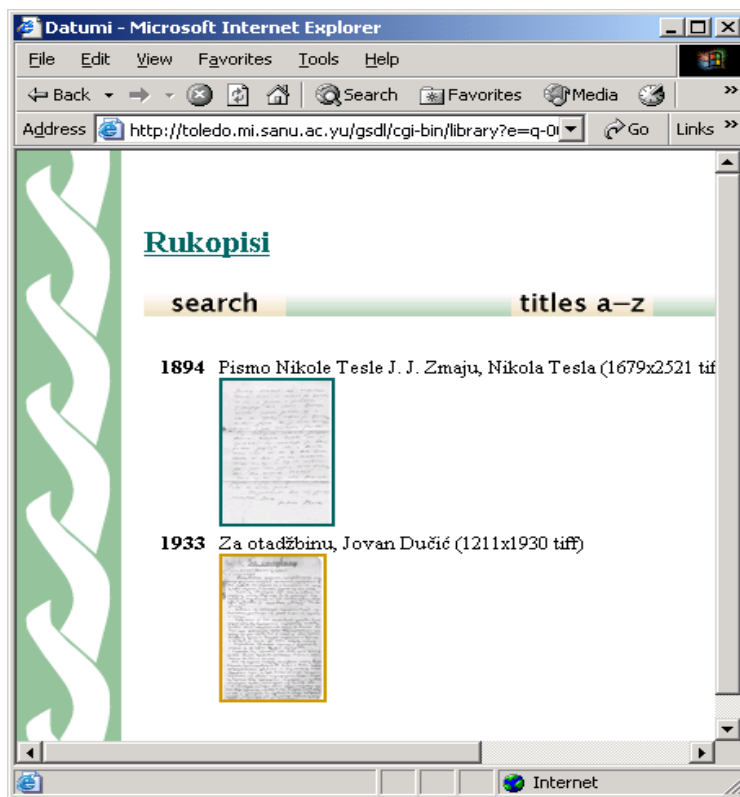
Zbirka *Zenit* sadrži samo prvi broj časopisa iz 1921. godine. Do njega se može doći pregledom liste srede po datumima, a pretraživanje uključuje autore članaka, naslove članaka, same članke i čitav časopis. U skladu sa svojom internacionalnom

orijentacijom, Zenit sadrži više jezika (srpskohrvatski, nemački, francuski) i više pisama (ćirilčno i latinično) i sve je u potpunosti podržano. Pretraživanje je ograničeno na pismo u kome se pojavljuje traženi pojam. Puni tekst (HTML) je takođe strukturiran, pa se može videti pregled članaka s naslovima i autorima. Ovde bi mogao da se dogradi PDF format. Na slici 2 prikazani su rezultati pretraživanja u kolekciji Zenit. Tražena je reč na ćirilici i pronađena u jednom dokumentu, koji se može pregledati.



Slika 2. Pretraživanje zbirke *Zenit* ćirilčnim pismom

Zbirkaa *Rukopisi* sadrži dva rukopisa do kojih se može doći iz liste sređene po datumima nastanka, kao i iz abecedne liste naslova. Originalni dokumenti su prilično veliki i nalaze se u pozadini, mogu se po potrebi pogledati ili preuzeti, a zastupljeni su minijaturnim slikama. Pretraživanje je moguće po ključnim rečima, po autorima i naslovima. Na slici 3 dat je prikaz minijaturnih slika rukopisa, sređen prema godini datiranja.



Slika 3. Pregled dokumenata u zbirci *Rukopisi*

Testiranje ovog softvera dalo je zadovoljavajuće rezultate za kreiranje digitalnih zbirki. Treba nastaviti s testiranjem za ostale vidove građe, prevesti interfejs na srpski, doterati grafički dizajn. Paralelno treba razmišljati o razvoju sopstvenog softvera, možda baš na iskustvima *Greenstone-a*.

Izvori

[1] How to Build a Digital Library / Ian H. Witten, David Bainbridge. – San Francisco: Morgan Kaufmann Publishers, 2003.

[2] Internet lokacija <http://www.greenstone.org>

tamara@nbs.bg.ac.yu

TESTING OF *GREENSTONE*, THE SOFTWARE FOR DIGITAL COLLECTIONS

National Library of Serbia and National Center for Digitization cooperated for a first time on the concrete project, on testing of open-sourced software for digital libraries, named *GREENSTONE*. As a result, there are a few digital collections made in *GREENSTONE*, available for browsing on the web location: <http://toledo.mi.sanu.ac.yu/gsdj/cgi-bin/library>.

Keywords: *GREENSTONE*, digital collections, digital documents