

DIGITALIZACIJA NOVINA „BEČEJSKI DANI“**DIGITIZATION OF THE NEWSPAPER „BEČEJSKI DANI“**Jelena Janković Ćuk, Željko Zeljković, *Fakultet tehničkih nauka, Novi Sad***Oblast – GRAFIČKO INŽENJERSTVO I DIZAJN**

Kratak sadržaj – Postupak digitalizacije je obiman posao, vođen projektom, koji se sastoji od upoznavanja i pravilne pripreme arhivske građe, transformacije u bit mapirane slike uz pomoć skenera i fotoaparata, preuređivanja i sređivanja podataka uz pomoć OCR softvera, kreiranja metapodataka i upravljanju baze, i na samom kraju kreiranje web aplikacije.

Ključne reči: novine, digitalizacija, štampani mediji, digitalni mediji, mikrofilm, mikrofilmovanje, OCR, skeneri, meta podaci, Dublin Core, XML, XSL.

Abstract – The procedure of digitalization of images is complex and usually, it's a case based procedure depending on the nature of the particular project: from making and preparing the archive information about the image, proceeding with scanning and digital preparation of image then converting in bitmap format by using OCR software, adding the metadata and finally preparing it for usage in web applications.

Keywords: newspapers, digitization, print media, digital media, microfilm, microfilming, OCR, scanners, metadata, Dublin Core, XML, XSL.

1. UVOD

Digitalizacija je pretvaranje analognoga ulaza u digitalni oblik. Cilj je da se pretvaranjem ne gubi na kvalitetu ulazne građe [1]. Digitalizacija predstavlja jedan detaljno planiran i skup projekat, koji treba voditi u skladu sa poznatim načelima i prioritetima i prema utvrđenim pravilima. U radu su analizirani metapodaci u bibliotekarstvu, na bazi čega se postavio koncept metapodataka u novinama.

2. DIGITALIZACIJA NOVINA

Prednost digitalizovanih novina su: mediji se mogu jednostavno i brzo umnažavati, a da se ne ošteti original, veliki broj korisnika imaju pristupa građi bez obzira na vremensko i geografsko ograničenje, sadržaj se prilikom korišćenja ne oštećuje, digitalni sadržaj je moguće pretraživati i menjati. Nedostaci su: digitalna zaštita, digitalizacijom se slike pretvaraju u bitove koji se zatim moraju rekonstruirati da bi se slike videle, a pri tom se senoviti prelazi na slikama mogu pretvoriti u oštre kontraste, za čitanje električnog dokumenta je potrebno imati dodatnu opremu, velika finansijska sredstva su potrebna za digitalizaciju.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Željko Zeljković, docent.

2.1. Elektronski mediji prednosti i mane u odnosu na štampane medije

Osnovna razlika između štampanih i elektronskih medija počiva na **dostupnosti i pokrivenosti**. Razlika između štampanih medija i elektronskih medija data je tabeli 1.

Tabela 1. Razlike štampanih i elektronskih medija

Osnova za upoređivanje	Štampani mediji	Elektronski mediji
Značenje	Medij koji dostavlja vesti i informacije putem štampanih publikacija.	Medij koji stvara, isporučuje i pristupa, vestima i informacijama putem elektronskih uređaja.
Pismenost	Treba biti pismen da bi pročitao date informacije.	Nije primarni zahtev, svako može gledati i čuti pružene informacije.
Rok	Rok postoji u vezi sa prikupljanjem vesti.	Vesti se mogu ažurirati stalno.
Diskusija	Nije moguća	Moguća
Pokrivenost	Manja, zavisi od mnogo faktora.	Efikasan u širenju vesti na nacionalnom i međunarodnom nivou.
Jezik	Pogodan za gledaoce	Pogodan za slušaoce
Ažuriranje	Periodično	Često
Dostupnost	Korisnik je slobodan da čita u svakom trenutku iz novina, šta god želi da čita i može ih nositi bilo gde.	Moguće je preotavati i gledati na različitim mestima ali za to su nam potrebni savremeni uređaji i Internet.
Kampanje	Relativno lakši oblik dostupnosti javnosti za kampanje	Trajanje je ograničeno na emitovanje.
Trajanje	Omogućavaju čitaocu uvid u članke u bilo kom trenutku u budućnosti	Elektronski zavise od stabilnosti i kapaciteta servera i baze
Prihod	Veliki rashodi u proizvodnji i distribuciji sadržaja	Relativno više prihoda
Inovativnost	U većoj meri je urađen maksimum što se tiče mogućnosti	Inovativniji oblik medija koji uključuje filmove, animaciju
Hendikep	Teško dostupan kod osoba sa oštećenjem vida	Bolje radi za osobe sa oštećenjima sluha i vida.

Štampani mediji se mogu opisati kao sredstvo masovne komunikacije, koje se koristi za informisanje šire javnosti putem štampanih publikacija, kao što su novine, časopisi, knjige itd. Elektronski mediji su noviji oblik masovnih medija u kojem se elektronski uređaji koriste za kreiranje i širenje vesti i informacija.

2.2. Tehnički zahtevi i sprovođenje digitalizacije

Digitalna slika je elektronska fotografija mapirana u obliku niza elemenata slike (piksela) i složena u skladu sa prethodno definisanim odnosom horizontala i vertikala [2].

Izvorna dokumenta se uz pomoć skenera ili fotoaparata prevode u bitmapirane slike. Skenirani dokument se čuva kao digitalni zapis u slike bez kompresije podataka (npr. formata TIFF) ili sa kompresijom (formata JPEG), i takve se slike mogu naknadno obraditi na računaru, pri čemu se može povećati ili smanjiti njihova svetlina, kontrast, intenzitet boje, oštrina i slično. Razlozi zbog kojih se mora vršiti kompresija slike: veliki zahtevi za memorijskim prostorom, spori uređaji koji nisu u mogućnosti da prikazuju nekompresovanu sliku u realnom vremenu, propusni opseg mreže koji ne dozvoljava prenos slike u realnom vremenu.

3. MIKROFILMOVANJE

Mikrofilmovanje je jedan od najefikasnijih, ekonomičnijih i jednostavnijih skladištenja velikih količina informacija na različitim tipovima filmova: srebrohalogenoidni, vizualni i dijazo film. Prednosti mikrofilmovanja: dokument na filmu je identična kopija originala, moguće je sačuvati dokumenta velikog formata u ukupnoj perspektivi slike, povezani sa kompatibilni izlaznim uređajima imaju mogućnost štampanja, faksiranja i slanja e-poštom, najjeftiniji način dobijanja arhivskih informacija, dostupnost nije ograničena elektronskim uređajima, mnogo informacija se može sačuvati na manjem prostoru nego što zauzima štampani materijal, može da se skenira i na taj način obezbedi identična kopija koja se trajno može čuvati.

Nedostaci mikrofilma su: korišćenje filma je ograničeno na jednu osobu, korisnici moraju imati uređaje za čitanje, za umnožavanje je potrebno imati odgovarajući štampač ili kopir, stranice se moraju čitati redom, nema mogućnost pretraživanja određene informacije, zapisi na mikrofilmu se ne mogu ažurirati bez upotrebe mikrofišnih tehnologija, trajnost i upotrebljivost filma regulisana je mnogim faktorima (kvalitet filma, pravilna kontrola obrade, odgovarajući uslovi skladištenja i rukovanje).

4. OCR SOFTVERI

Da bismo izvukli i preuredili podatke iz skeniranih dokumenata, slika sa fotoaparata ili PDF-ova, potreban je OCR softver koji će omogućiti uređivanje sadržaja originalnog dokumenta. Bitno je da OCR softver ima podršku rečnika na jeziku na kojem je napisan izvorni dokument. Princip rada OCR softvera: program analizira strukturu slike dokumenta, nakon čega program deli stranicu na elemente kao što su blokovi tekstva, tabela, slika, itd; Linije su podeljene u reči, a zatim u znakove. Kada se skenirane slike izdvoje, program ih upoređuje sa skupom slika uzoraka. On unapređuje brojne hipoteze o mogućnostima skeniranih slika. Na osnovu ovih hipoteza program analizira različite varijante razbijanja linija u reči i reči u znakove. Nakon obrade ogromnog broja takvih verovatnoća hipoteza, program konačno donosi odluku, predstavljajući prepoznatljiv tekst.

5. SKENERI

Vrste skenera prema načinu rada, mogu biti: ručni (handheld), ravni (flatbed), sa uvlakačem (sheetfed), rotacioni (drum). Vrste skenera prema uzorcima: papir (tekst, slika), dijapozitivi, mikrofilm, negativi (pomoću šape za skeniranje negativa), 3D skeneri (predmeti). Osnovne karakteristike skenera su: brzina skeniranja (zavisi od klase i generacije skenera), maksimalna rezolucija [3].

5.1. Skeneri velikih formata za digitalizaciju

Mogućnosti skenera velikih formata: skeniranje u boji ili crno/belo, veoma su brzi i jednostavni za korišćenje, Dual DSLR sistem fotoaparata/kamera, skeniranje bez zakrivljenosti listova zahvaljujući staklenoj ploči, integrisan softver za skeniranje, postprodukciju i kontrolu radnog procesa (workflow), sto za postavljanje uređaja: kompaktan, prenosiv, jednostavno postavljanje, intuitivan rad na stonom uređaju, skenirane podatke može da smešta direktno na USB, može biti povezan na štampač, a može imati izlaz i za bilo koji mrežni drajver. Neke od posebnih funkcija sa kojima smo se susreli u istraživanju skenera velikih formata su: Scan2Pad opcija koja omogućava bežičnu konekciju sa skenerom direktno na mobilne uređaje kao što su iPad ili Android tablet, mogućnost da sa skenerom može direktno upravljati sa iPada ili tableta, a skenirane slike su odmah dostupne na mobilnim uređajima bilo da se skladište ili dalje obrađuju; patentirani mehanizam za okretanje "Bionički Prst", ali poneki skeneri imaju i mogućnost da skeniraju obe strane dokumenta u samo jednom prolazu širine do 915 mm.

Pri konverziji materijala je veoma bitno:

- Odabrati skenere koji se mogu prilagoditi fizičkim dimenzijama izvornog dokumenta, vrsti medijuma, rasponu detalja, tonova i boja na dokumentima, kao i fizičkom stanju dokumenata
- Odabrati rezoluciju, bit dubinu koja je dovoljna da prenese najfinije bitne detalje u novinama koje skeniramo i da veoma pažljivo primenjujemo postupke poboljšavanja novina.
- Koristiti standardne tehnike kompresije "bez gubitka kvaliteta" za matične tj. master datoteke
- Pokušati postići balans između vizuelnog kvaliteta prihvatljivog korisnicima i veličine fajla kojem računar može pristupiti sa prihvatljivim zakašnjenjem.

6. METAPODACI

Metapodaci su strukturirane informacije koje opisuju, objašnjavaju, lociraju ili na drugi način čine lakšim pronalaženje, korišćenje ili upravljanje nekim izvorom informacija. Metapodaci se često definišu kao podaci o podacima ili informacije o informacijama. [4] Metapodaci mogu biti opisni, strukturalni i administrativni.

Metapodaci mogu biti ugrađeni u digitalni objekat, a mogu se čuvati i odvojeno. Metapodaci su često ugrađeni u HTML dokumente i u zaglavlja slikovnih datoteka. Čuvanjem metapodataka zajedno sa objektom, osigurava se da oni neće biti izgubljeni, uz to se izbegavaju problemi povezivanja podataka i metapodataka i

dobija se na sigurnosti da će se metapodaci i objekat ažurirati zajedno. [5] Loše osobine metapodataka: kreiranje metapodataka je skupo i dugotrajno, metapodaci su komplikovani (naročito za krajnje korisnike), metapodaci su subjektivni i zavise od konteksta, nema jasne granice do kog nivoa detalja su metapodaci korisni, metapodaci su beskorisni (imajući u vidu fulltext pretraživača kao što je Google). [6]

6.1. Meta podaci u bibliotekarstvu

Formati za mašinski čitljivo beleženje bibliografskih podataka: MARC21 i UNIMARC. Dva najveća biblioteka sistema u Srbiji, COBISS i BISIS, rade na UNIMARC platformi.

Uloga metapodataka u bibliotekarstvu je da se pronađe izvor podataka prema relevantnim kriterijumima; da se organizacije i upravlja različitim arhivama; da se omogući razmena informacija i sadržaja između različitih sistema, struktura podataka i interfejsa uz minimalne gubitke na sadržaju i funkcionalnosti; zatim digitalna identifikacija (preko trajnih identifikatora kao što su URL adrese ili DOI brojevi); i na samom kraju arhiviranje i zaštita digitalnih informacija. Šeme za metapodatke: Dablinsko jezgro (Dublin Core), TEI (Text Encoding Initiative), METS (Metadata Encoding and Transmission Standard), MODS (Metadata Object Description Schema), LOM (Learning Object Metadata).

6.2. Dublin Core

Osnovne karakteristike Dublin Core su: jednostavan, prilagodljiv, međunarodan, proširiv, sami elementi su opcioni, ponovljivi, mogu se pojavljivati u bilo kom redosledu, sadrži šifarnik i njegova primena se preporučuje.

Primena Dublin Core Metadata Elementa u biblioteci:

- Služi kao format razmene između različitih sistema koristeći različite standarde / formate metapodataka
- Koristiti se za prikupljanje metapodataka iz izvora podataka unutar i izvan bibliotečkog domena
- Podržava jednostavno kreiranje zapisa iz kataloga biblioteka za resurse u različitim sistemima
- Izlagati MARC podatke drugim zajednicama (kroz konverziju u DC), kako bi se omogućilo pribavljanje metapodataka otkrivanja resursa od stvaralaca koje nisu biblioteke koje koriste DC.

Pri odabiru šeme potrebno je voditi računa da:

- Datoteke skeniranih slika budu organizovane tako da se poštuje hijerarhija diska koja se logično poklapa sa fizičkom organizacijom dokumenata.
- Imenujemo skenirane datoteke na strogo kontrolisani način koji odražava njihove logičke odnose.
- Opišimo datoteke skeniranih slika interno (pomoću zaglavlja slikovne datoteke) i eksterno (pomoću povezanih datoteka sa deskriptivnim metapodacima).
- Digitalizovane novine moraju da se zaštite, i postoji potreba da povrate uložena sredstva, pa može da razmišlja u smeru da naplaćuje uslugu fotokopiranja digitalizovanih novina, downloada određenih pdf-ova i slika, kroz mesečne ili dnevne članarine.

7. XML, XSL

Rezultati rada

Tabela 2. Primer Dublin Core metapodataka za novinski članak: Bečejski dani br.1, 28.05.1994. godine, Članak 1

dc.title	Vlasnička transformacija Fadipa
dc.subject	ekonomija, fadip, nikola dolinka, deonice, zaposleni, penzioneri, savo jelaca, drustveni kapital...
dc.description	Holding Fadip je izvršio dodatnu transformaciju društvenog kapitala u privatnu, čime je izazvao revolt 200, što bivših, što penzionisanih radnika ovog preduzeća...
dc.creator	K. K. Šanta
dc.publisher	Zoran Subotički, CANNON M
dc.date	19940528
dc.type	text
dc.format	text/html
dc.identifer	http://becejskidani.co.rs/enovine/br1_str1.pdf.
dc.source	Novine „Bečejski dani”
dc.language	sr
dc.coverage	
dc.rights	Sva prava zadržava Medium Doo, Bečejski dani

Prikaz XML šeme, za izdanje br.1 od 28.maja 1994. godine:

```
<?xml version="1.0" encoding="UTF-16"?>
<!DOCTYPE nbd SYSTEM »nbd.dtd«>
<!?xmlstylesheet type="text/xsl" href="»nbd_to_html.xslt«?>
<nbd>
<p>Arhiva digitalizovanih novina »Bečejski dani«, sadrži preko 2000 izdanja. Prvo digitalizovano izdanje je iz 6. januara 1962. godine. Pretragu je moguće izvršiti na osnovu zadatog kriterijuma, u svim segmentima jednog članka, naslovu, autoru, predmetu članka, izdavaču, opciono na osnovu tipa dokumenta i jeziku na kojem je napisan članak . Takođe pretragu možemo ograničiti i na konkretno neki vremenski period za koji nas zanima određeni kriterijum.</p>
<clanak>
<title> Endre Husag predsednik opštine </title>
<creator>K. K. Šanta</creator>
<publisher>Zoran Subotički, CANNON M</publisher>
<subject>predsednik opštine becej, endre husag, lokalna samouprava Becej,intervju, politika</subject>
<description>Intervju sa predsednikom opštine Becej Endre Husagom o njegovim ingerencijama i mogućnostima po pitanju kvalitetnijeg organizovanja života u bečejskoj opštini, o tome koji su najozbiljniji problemi u bečejskoj opštini, o vodosnabdevanju, o međustranačkoj saradnji na lokalnom nivou.</description>
<date>19940528</date>
<type>text/jpg</type>
<language>sr</language>
<identifer>http://becejskidani.co.rs/enovine/br1_str4.pdf/ </identifier>
<rights>Sva prava zadržava Medium doo, Bečejski dani</rights>
</clanak> </nbd>
```

Prikaz DTD šeme:

```
<?xml version="1.0" encoding="UTF8" ?>
<!ELEMENT nbd (p, clanak+)>
<!ELEMENT p (#PCDATA)>
<!ELEMENT clanak (title, creator, publisher, subject,
description, date, type, language, identifier,
rights)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT creator (#PCDATA)>
<!ELEMENT publisher (#PCDATA)>
<!ELEMENT subject (#PCDATA)>
<!ELEMENT description (#PCDATA)>
<!ELEMENT date (#PCDATA)>
<!ELEMENT type (#PCDATA)>
<!ELEMENT language (#PCDATA)>
<!ELEMENT identifier (#PCDATA)>
<!ELEMENT rights (#PCDATA)>
```

Prikaz XML u web pretraživaču na slikama 1 i 2:

Arhiva digitalizovanih novina "Bečejski dani", sadrži preko 2000 izdanja. Prvo digitalizovano izdanje je iz 6. januara 1962. godine. Pretragu je moguće izvršiti na osnovu zadatog kriterijuma, u svim segmentima jednog članka: naslovu, autoru, predmetu članka, izdavaču, opciono na osnovu tipa dokumenta i jeziku na kojem je napisan članak.

Detaljnija pretraga:
 naslov autor predmetu članka izdavaču kompletno sve
Tip dokumenta:
 text jpg text/jpg
Jezik članka:
 srb hu

Slika 1. Prikaz pretrage

Bečejska privreda na Novosadskom sajmu

Autor: V. Bekvalac

Izdavač: Zoran Subotički, CANNON - M

Ključne reči: privreda Bečej, Pik-Bečej, Sojaprotein, Karbodioksid, Fadip, međunarodni sajam, Slobodan Milošević, Mirko Marjanović, Milutin Stojković

Kratki opis: Među 1.600 izlagača na 61. međunarodnom sajmu u Novom Sadu, izlagača su i četiri bečejska preduzeća: Pik-Bečej, Sojaprotein, Karbodioksid i Fadip i opšta je ocena da su svi bili zapaženi. Najveće interesovanje je bilo na štandovima kombinata Pika-Bečej. Uz brojna priznanja oni su imali priliku da na svom štandovima dočekaju i pozdrave predsednika Republike Srbije Slobodana Miloševića, Mirka Marjanovića, predsednika Vlade Srbije, dr Milutina Stojkovića, predsednika Skupštine Vojvodine.

Datum izdanja: 1994-05-28

Tip dokumenta: text

Jezik članka: sr

Link: http://becejskidani.co.rs/enovine/br1_str6.pdf

Prava: Sva prava zadržava Medium doo, Bečejski dani

Slika 2. Prikaz članka

8. ZAKLJUČAK

Digitalizacija našeg kulturnog nasleđa spaja različite sektore globalne zajednice na sasvim novi način. Usled velikog korišćenja Interneta, korisničke grupe bibliotečkih struktura su se transformisale. Izdavači novina, integrišu štampana i online izdanja da bi doprli do šireg kruga čitalaca.

Interaktivne tehnike Interneta pružaju bibliotekama i arhivima novu priliku da razviju globalnu zajednicu korisnika, pa tako i novine imaju potrebu da prate ovaj trend kako bi opstale. Stara arhiva ima potrebu za digitalizovanjem, a nova izdanja imaju potrebu za pravilnim očuvanjem i organizacijom.

Razloga za digitalizovanjem novina iz 60-ih godina XX veka, koji su u našoj arhivi je pregršt:

- povećali bi pristup objavljenim informacijama iz tog perioda;
- poboljšali bi usluge rastućoj grupi korisnika obezbeđivanjem kvalitetnijeg pristupa izvorima informacija za potrebe informisanja;
- smanjilo bi se rukovanje i korišćenje osetljivog i često upotrebljavanog izvornog materijala,
- kreirale bi se rezervne odnosno sigurnosne kopije;
- razvili bi se zajednički resursi, uspostavilo partnerstvo sa drugim institucijama u cilju kreiranja virtuelnih zbirki i povećanog globalnog pristupa;
- uspostavilo partnerstvo sa drugim institucijama i ostvario profit.

Krajnji izazov za redakcije je kreiranje WEB aplikacije i za to je potrebno da zaposlene koji poseduju izvesne tehničke, upravljačke i uređivačke veštine organizovati u web tim. Neophodno je usvojiti tehničke smernice da bi osigurali trajno visoke standarde u izradi i upravljanju web sajtom.

9. LITERATURA

- [1] S. Lukačević, J. Buljević, S. Mokriš, "Formati digitalnih sadržaja za čuvanje i zaštitu novina u knjižnicama", Str. 104, 2015.
- [2] IFLA, "Smjernice za projekte digitalizacije", Str. 29, 2002.
- [3] Elektrotehnički fakultet u Banja Luci, katedra za opštu elektrotehniku, [Online] Dostupno na: <http://dsp.etfb.net/multimediji/2015/07%20GI%20Kompresija%20slike.pdf>; [Pristupljeno: 27. marta 2019. godine]
- [4] L. Hodž, 2004. [Online] Dostupno na: http://poincare.matf.bg.ac.rs/~cvetana/Nastava/Materijal/GejlHodz_metapodaci.pdf; [Pristupljeno: 5. novembra 2019. godine]
- [5] N. Dakić, J. Andonovski, 2016, *Linked Open Data u Bibframe*, [Online] Dostupno na: <https://www.unilib.rs/2016/03>; [Pristupljeno: 12. novembra 2019. godine]
- [6] Milosavljević B., Zaric M, Zeljković Ž., *Elektronsko izdavaštvo materijal sa predavanja*, 2018.

Adresa autora za kontakt:

MSc Jelena Janković Ćuk: jeckojeckic@gmail.com

Dr. Željko Zeljković: zeljkoz@uns.ac.rs

Grafičko inženjerstvo i dizajn,

Fakultet Tehničkih Nauka, Novi Sad