

**Denis Despot
Domagoj Tominac**

Filozofski fakultet Sveučilišta J.J. Strossmayera u Osijeku
Odsjek za informatologiju

DIGITALIZACIJA OSJEČKE BIBLIOGRAFIJE MARIJE MALBAŠE

Tijekom studentske prakse koju smo provodili u Muzeju Slavonije u Osijeku upoznali smo se pobliže s raznim problemima u korištenju muzejske građe, problemima pretraživosti, osjetljivošću knjige. Koliko god pisana riječ bila jaka, njen tradicionalni medij je vrlo podložan zubu vremena i propadanju. Zahvaljujući prof. Marini Vinaj došli sam na zamisao o digitalizaciji jedne značajnije skupine građe u sklopu studentske prakse, a ujedno i na korist Muzeju. Izbor pao na Osječku bibliografiju Marije Malbaše. Osim prevencije propadanja, koja zapravo ne prijeti ovom djelu s obzirom na godinu izdanja (1981. godina), ono što je najveća prednost kod digitalizacije ovoga djela olakšano korištenje u digitalnom obliku. Budući je riječ o bibliografiji, a u digitalnom obliku je pretraživa i po jednom elementu, puno je lakše pronaći traženu jedinicu. U ovom radu ćemo pokušati navesti osnove procesa digitalizacije, nekih već zahuktalih inicijativa te osvrnuti se na djelo koje smo prenijeli u digitalni oblik, kao i samu metodologiju.

DIGITALIZACIJA

Digitalizacija kao postupak prebacivanja dokumenata (informacija, znanja) iz analognog u digitalni oblik, uz dodatnu obradu i arhiviranje danas je postupak pomoću kojeg različiti dokumenti postaju javno dostupni korisnicima, a imaju i svrhu očuvanja izvornih dokumenata¹. U knjižničarstvu digitalizacijom najčešće podrazumijevamo prijenos analogna građe u digitalni format². Pod analognom građom podrazumijevamo građu u tradicionalnoj tehnologiji (rukopise, knjige, časopise, novine, mikrofilm, grafiku, razglednice itd.) dok se pod digitalnom građom podrazumijeva sva grada rađena u elektroničkoj tehnologiji (medij - fizički nositelj informacije je CD, DVD, disketa, magnetska traka...).

¹ *The Oxford text archive: Creating and documenting electronic texts. URL:* <http://ota.oucs.ox.ac.uk/documents/creating/cdet/chap3.html#0> (1.9.2010)

² *Batnožić, Slaven. Digitalizacija starih izdanja.// Vrijenac, 171(2000) URL:* <http://www.matica.hr/Vrijenac/Vij171.nsf/>
AllWebDocs/Digitalizacijastarihizdanja (1.9.2010)

INICIJATIVE

U digitalizaciji tiskanih knjiga i njihovom raspačavanju putem Interneta projekti *Gutenberg*³ i *netLibrary.com*⁴ mogu se smatrati međašima, iako su kasnije mnogi slijedili njihov primjer. Nastali su u vremenskom razmaku od punih 28 godina i predstavljaju dva pristupa digitalizaciji tiskanih knjiga. Dok *Gutenberg* funkcionira na dobrovoljnoj osnovi i digitalizira naslove nezaštićene autorskim pravom, *netLibrary.com* se upustio u digitalizaciju naslova zaštićenih autorskim pravom, našao rješenja za poslovanje elektroničkom knjigom, osigurao autorima zaštitu i postavio poslovanje na komercijalnu osnovu, koliko su bili uspješni ilustrira činjenica da je početkom 2010. godine EBSCO⁵ kupio *netLibrary* i počeo, između ostalih i uslugu nadopunjavanja knjižničnih fondova prodajom digitaliziranih sadržaja⁶.

Postoje i korpusi digitaliziranih tekstova koje razni centri stavljuju na mrežu, često ih nazivajući i arhivima. Zapravo nema prepreke da sve ustanove (centri, knjižnice) digitaliziraju rukopisne i tiskane tekstove koji predstavljaju javno dobro te činom digitalizacije i stavljanja na mrežu postaju njihovi nakladnici. Poznate zbirke elektroničkih tekstova su npr. *Oxford Text Archive*⁷ i *Scholars' Lab - University of Virginia*^{8 9}. Uversitiy of Virginia se uz još nekoliko velikih knjižnica¹⁰ uključila u Googleov projekt masovne digitalizacije o kojemu nešto više u dalnjem tijeku teksta.

GOOGLE BOOKS

Projekt koji smo izdvojili smatramo jednim od najvažnijih inicijativa koja omogućuje pregledavanje knjiga u digitaliziranom obliku, a to je *Google Books*¹¹.

Sve je počelo 1996. godine dok su suosnivači Googlea Sergey Brin i Larry Page još bili studenti računalnih znanosti i radili na istraživačkom projektu koji je podržavao Stanford Digital Library Technologies Project. Njihov je cilj bio pokretanje rada digitalnih knjižnica, a misao vodilja im je bila da u svijetu budućnosti u kojem su velike zbirke knjiga digitalizirane, ljudi koriste „web indeksiranje“ kako bi indeksirali sadržaj knjiga i analizirali veze među njima utvrđujući relevantnost i korisnost bilo koje knjige praćenjem broja i kvalitete citata iz drugih knjiga. 2002. godine skupina Googleovih zaposleni-

³ Free eBooks by Project Gutenberg. URL: <http://www.gutenberg.org/> (1.9.2010)

⁴ LibraryNet. URL: <http://netlibrary.com/> (1.9.2010)

⁵ EBSCO Publishing. URL: <http://www.ebscohost.com/> (1.9.2010)

⁶ Research Information. URL: http://www.researchinformation.info/news/news_story.php?news_id=588 (1.9.2010)

⁷ The Oxford text archive. <http://ota.ahds.ac.uk/> (1.9.2010)

⁸ Scholars' Lab. URL: <http://www2.lib.virginia.edu/scholarslab/> (1.9.2010)

⁹ University of Virginia Library. <http://lib.virginia.edu/> (1.9.2010)

¹⁰ Mass digitization of books and digitized libraries. <http://massdigitization.com/> (1.9.2010)

¹¹ Google Books. URL: <http://books.google.com/> (1.9.2010)

ka službeno pokreće projekt „books“. Počinju razgovarati sa stručnjacima o onome što ih očekuje, počevši od jednostavnog ali bitnog pitanja: koliko bi vremena bilo potrebno da se skeniraju sve knjige na svijetu? Naravno 2002. godine to nitko nije znao, a nitko sa sigurnošću ne može tvrditi niti danas. Larry Page to je odlučio iskušati sam, tako on i Marissa Mayer, jedna od prvih rukovoditeljica projekta, pomoću metronoma održavaju ritam tijekom metodičkog okretanja listova knjige od 300 stranica. Do kraja knjige trebalo im je 40 minuta¹². Tijekom 2003. godine tim koji se bavi ovim projektom smišlja tehnike skeniranja koje ne uništavaju knjige. Nakon mnogih pokusa, tim razvija metodu skeniranja koja je mnogo blaža od trenutno uobičajenih procesa velikih brzina. U isto vrijeme softverski inženjeri napreduju prema rješavanju zahtjevnih tehničkih problema s kojima se susreću pri procesiranju informacija iz knjiga koje sadržavaju neuobičajene veličine znakova, neobične fontove ili druge neočekivane posebnosti – na 430 različitih jezika. 2004. godine nakon posjete Bodleiani na Oxfordskom Sveučilištu dolazi do službenog partnerstva za digitalizaciju neprocjenjive knjižnične zbirke koja broji više od milijun knjiga iz 19. stoljeća. U međuvremenu, serija pregovora s nekim od najvećih svjetskih izdavača počinje donositi plodove, tako u listopadu 2004. godine Larry i Sergey najavljaju „Google Print“ na Sajmu knjiga u Frankfurtu. Izdavači koji su se prvi pridružili programu bili su: Blackwell, Cambridge University Press, University of Chicago Press, Houghton Mifflin, Hyperion, McGraw-Hill, Oxford University Press, Pearson, Penguin, Perseus, Princeton University Press, Springer, Taylor & Francis, Thomson Delmar i Warner Books. Iste godine Google najavljuje početak projekta „Google Print“ koji je omogućilo partnerstvo s Harvardom, Michiganskim sveučilištem, Javnom knjižnicom New Yorka, Oxfordom i Stanfordom. Procijenjeno je da zbirke svih ovih knjižnica zajedno imaju više od 15 milijuna svezaka. Tek 2005. godine, Google se širi na 8 europskih zemalja i mijenja naziv „Google Print“ u „Google Books“, naziv koji puno bolje dočarava prirodu usluge.

Google books je još uvijek u beta verziji ali to je tipična Googleova politika zadržavanja naziva beta u imenu svojih usluga dugi niz godina. Osobno smatram kako je upravo ovaj projekt najperspektivniji od svih trenutno aktivnih inicijativa poput *Europeane*, *netLibrary*, *Gutenberg*, *openLibrary* i ostalih upravo zbog kapitala, ljudi, iskustva, poslovičnoj odlučnosti i inovativnosti koju Google ima.

OCR

Optičko prepoznavanje znakova (Optical Character Recognition) je dizajniran kako bi prevodio sliku otisnutog teksta s papira, koju obično učitamo

¹² Google Books. URL: <http://books.google.com/googlebooks/about.html> (1.9.2010)

sa skenera u tekst koji se može uređivati računalom, ili prevodi sliku sa znakovima u standardnu kodnu shemu predstavljajući ih u ASCII ili UNICODE kodu.¹³

OCR je izvorno započeo kao polje u istraživanju umjetne inteligencije i strojnog gledanja. Na početku su postojale dvije odvojene grane: optičko raspoznavanje znakova uz pomoć korištenja optičkih pomagala kao što su ogledalo i leće; te digitalno raspoznavanje znakova uz korištenje skenera i kompjuterskih algoritama. Pošto je većina optičkih tehnika zastarjela ili nisu tako često korištene, termin optičkog raspoznavanja znakova postao je uobičajen, te se ujedno koristi i za digitalno raspoznavanje znakova.

Za razliku od ranijih softvera danas imamo „inteligentne“ OCR softvere koji mogu prepoznati većinu fontova sa visokim stupnjem točnosti. Neki od njih su sposobni približno reproducirati originalno skeniranu stranicu što uključuje slike, stupce i druge netekstualne komponente.¹⁴

Program koji sam ja koristio za digitalizaciju Osječke bibliografije je „ABBYY FineReader“¹⁵. Od ostalih vrijedi navesti „Top Image Systems“¹⁶ koji su jedni od pionira razvoja OCR programa ali su prestali izdavati svoj svojevremeno solidan softver i bazirali se na pružanje svakojakih usluga vezanih za strojno čitanje; „CuneiForm“¹⁷ koji iako je počeo kao profesionalni sustav za veće tvrtke optimiziran za rad s velikim obimom materijala danas ga nalazimo u besplatnoj inačici otvorenog koda, te „Tesseract“¹⁸ izvorno HP-ov program koji od 1995. godine razvija Google no prilično je zahtjevan (traži .TIFF datoteke) i dojma smo da se na njemu ne radi mnogo te ga također možemo skinuti s interneta besplatno ali definitivno ne u inačici koju koriste Gooleovi inžinjeri.

POVIJEST TEHNOLOGIJE

Prvi sustavi za optičko prepoznavanje znakova pojavili su se 50-tih godina 20. stoljeća¹⁹, te su se temeljili na mehaničkim i optičkim metodama pribavljanja slike teksta kombiniranih s fotoćelijama. Razvojem digitalnih sklopova, a osobito zbog ostvarenja većih rezolucija fotoćelija razvijala se i tehnologija. Tijekom 60-ih i 70-ih godina prošlog stoljeća, tehnologija OCR-a je zaživjela u praksi u poštama, bolnicama i tvrtkama. United States Postal Service služi se OCR uređajima za razvrstavanje pošte od 1965.

¹³ Tekstilno-tehnološki fakultet: skeniranje. <http://racunala.ttf.hr/files/Skeniranje.pdf> (1.9.2010)

¹⁴ isto

¹⁵ ABBYY - OCR, ICR, OMR, Data Capture and Linguistic Software. <http://www.abbyy.com/> (1.9.2010)

¹⁶ Top Image Systems – Data Capture, Invoice Processing, Forms Processing, Document Capture & Classification. <http://www.topimagesystems.com/Page.asp?Par=1&id=8> (1.9.2010)

¹⁷ OCR CuneiForm. <http://www.cuneiForm.ru/eng/index.html> (1.9.2010)

¹⁸ Tesseract-OCR: Project Hosting on Google Code. <http://code.google.com/p/tesseract-ocr/> (1.9.2010)

¹⁹ Brief history of OCR: Encyclopedia II. http://www.experiencefestival.com/optical_character_recognition (1.9.2010)

Prva uporaba OCR-a u Europi bila je u British General Post Office-u²⁰. 1965. isplaniran je bankovni sustav koji koristi OCR tehnologiju po imenu National Giro. Taj je proces revolucionarno promijenio sustav plaćanja računa u Velikoj Britaniji. Canada Post koristi OCR tehnologiju od 1971.

Uz razvoj tehnologije OCR-a, izvedbe sustava su se nastojale poboljšati i na druge načine, jedan od njih je uvođenje specijaliziranih fontova za OCR. Koristenjem fontova poput OCR-A ili OCR-B²¹, razvijenih u Američkom nacionalnom institutu za standarde (ANSI²²) i Europskoj udruzi proizvođača računala (ECMA²³), sustavi za prepoznavanje znakova postizali su vrlo visok stupanj pouzdanosti. Spomenuti fontovi su posebno dizajnirani da računalnu obradu učine što jednostavnijom. Za razliku od uobičajenih fontova, u kojima je npr. slovo „i“ znatno uže od slova „m“, svi znakovi OCR fontova su jednakog širine. Time je segmentacija teksta na pojedine znakove znatno olakšana. Uz to, pri dizajnu fontova prilagođenih za OCR nastojalo se inače slične znakove npr. broj „1“ i slovo „l“ učiniti što različitijima. Iako je današnja OCR tehnologija znatno napredovala, fontovi OCR-A i OCR-B još uvijek se koriste. Razvojem hardvera i povećanjem brzine računala u zadnjih nekoliko desetljeća, nastao je čitav niz algoritama prepoznavanja i pred procesiranja koji se temelje na raspoznavanju uzorka, računalnom vidu i drugim tehnikama. Mnoge tehnike prepoznavanja znakova korijene vuku iz metoda umjetne inteligencije.

Danas je za većinu latiničnih fontova moguće raspoznavanje znakova uz vrlo visoku pouzdanost.

PREDNOSTI I NEDOSTACI

Prednosti digitalizacije su višestruke i u današnjem internetskom okruženju lako ih je primijetiti. Stavljanje građe na internet omogućava brzu i jednostavnu dostupnost korisnicima diljem svijeta i to bez ugrožavanja kvalitete kopije koju korisnik dobiva, pri umnažanju se naravno ne gubi na kvaliteti. Osim navedenih činjenica koje čine život lakšim krajnjem korisniku i knjižničarima, digitalna građa ne oštećuje se korištenjem, tj. kvalitetom s vremenom ne nestaje i ne umanjuje se uporabom kao što je to slučaj s konvencionalnom građom, stoga je i konvencionalni fond pošteden oštećenja nastalih fizičkom uporabom. Bitna prednost je vrlo jednostavno pretraživanje baze kao i svakog dokumenta, čak i na razini jednog slova kao i formuliranjem složenih upita.²⁴

²⁰ Brief history of OCR: Encyclopedia II. http://www.experiencefestival.com/a/Optical_character_recognition_-Brief_history_of_OCR/id/5354504 (1.9.2010)

²¹ Typographic Abbreviations Series #1: OCR. <http://myfonts.wordpress.com/2006/09/18/typographic-abbreviations-series-1-ocr/> (1.9.2010)

²² American National Standards Institute – ANSI. <http://www.ansi.org/> (1.9.2010)

²³ Ecma International. <http://www.ecma-international.org/> (1.9.2010)

²⁴ Smith, Abby. Why Digitize? Washington: Council on Library and Information Resources. 1999. URL: www.clir.org/pubs/reports/pub80-smith/pub80.pdf (1.9.2010)

Unatoč navedenim prednostima, digitalizacija ima i svoje nedostatke. Jedan od najčešćih nedostataka je teško rješivi problem pohrane, današnji mediji (čvrsti diskovi, CD i DVD mediji) još nisu dovoljno pouzdani i ispitani da bi se sa velikom sigurnošću na njih dugotrajno računalo²⁵ a za primjer, građa tiskana na beskiselinskom papiru i čuvana u optimalnim uvjetima (npr. u hladnom prostoru s malo vlage) može trajati više stotina godina²⁶. Taj problem se danas rješava migracijom, te kombiniranom pohranom na različite vrste medija. Očiti problem digitalne građe je potreba za digitalnom opremom jer nije poput papira čitljiv bez uporabe nekog vrsta računala²⁷, no ta oprema je danas toliko jeftina i sveprisutna da zbilja ne smije predstavljati ozbiljniji problem. Ipak, problem opreme neminovno povlači za sobom i pitanje koštanja korištenja i prije svega stvaranja digitalne građe, troškovi ovog procesa, još uvjek, nisu posve zanemarivi ali se napretkom tehnologije svakodnevno smanjuju. Ista stvar se može reći i za standarde pohrane i opisa koji su svakim danom sve detaljniji i optimizirani za korištenje.

ŠTO I ZAŠTO DIGITALIZIRATI?

Kako je digitalizacija još uvjek vrlo skup i spor proces, Ministarstvo kulture Republike Hrvatske je u sklopu nacionalnog projekta "Hrvatska kulturna baština"²⁸ izdalo naputke i smjernice za odabir građe za digitalizaciju koje sam uzeo kao relevantne za naše područje i uvjete u kojima se kao zemlja nalazimo. Iako sam tijekom istraživanja našao na više relevantnih stranica i sličnih smjernica u sklopu sveučilišta i organizacija poput Oxford-a²⁹ i UNESCO-a³⁰ njih ču samo spomenuti pošto srž odabira građe za digitalizaciju kao i sama svrha digitalizacije u svim smjernicama ostaje vrlo slična.

DIGITALIZACIJA RADI ZAŠTITE IZVORNIKA

Digitalizacijom se posredno ostvaruje zaštita građe jer se osiguranjem pristupa digitalnim preslikama smanjuje korištenje izvornika što utječe na njihovu bolju očuvanost. Dajući na korištenje digitalnu presliku, izvornik možemo trajno pohraniti u spremište s nadziranim uvjetima pohrane što je posebno značajno kod osjetljive građe poput starih rukopisa pisanih na

25 de Lusenet, Yola. *Long term access to the digital world: from journal to web*, ožujak 2004. URL: www.knaw.nl/ecpa/publ/ease_paper.pdf (1.9.2010)

26 de Lusenet, Yola. *Preservation of digital heritage: European commission on preservation and access*, ožujak 2002. URL: <http://www.knaw.nl/ecpa/PUBL/unesco.html> (1.9.2010)

27 Jurgens, Martin. *Preservation of ink jet hardcopies*, 27.8.1999. URL: <http://www.knaw.nl/ecpa/PUBL/jurgens.html> (1.9.2010)

28 Hrvatska kulturna baština - Nacionalni program digitalizacije. URL: <http://kulturna.hr/hr/0-nama> (1.9.2010)

29 AHDS Guides to good practice. URL: <http://ota.oucs.ox.ac.uk/documents/creating/cdet/> (1.9.2010)

30 de Lusenet, Yola. *Preservation of digital heritage: European commission on preservation and access*, ožujak 2002. URL: <http://www.knaw.nl/ecpa/PUBL/unesco.html> (1.9.2010)

različitim podlogama, novina, fotografskih negativa i sl.³¹ Ovo je segment u kojem je dugotrajan i skup proces digitalizacije opravdan. Poznavajući cijene uređaja za digitalizaciju npr. osjetljivih rukopisa apsolutno se slažem da je digitalizacija takvog izvornika skupa i dugotrajna no isto ne mora vrijediti za negative i novine gdje je proces digitalizacije višestruko jednostavniji.

U baštinskim ustanovama postupak digitalizacije nadopunjuje ostale postupke izrade zaštitnih snimaka/preslika građe radi zaštite, kao što je, primjerice, izrada mikrofilmskih preslika koja se još uvjek smatra iznimno pouzdanim sredstvom dugoročne zaštite.³² Ne ulazeći u detaljniju raspravu o prednostima i nedostacima izrade mikrofilmskih i digitalnih preslika, treba naglasiti da vrijednost digitalizacije u području zaštite građe raste s razvojem tehnologije koja osigurava sve bolje uvjete za dugoročnu pohranu i osiguranje čitljivosti računalnih datoteka.

DIGITALIZACIJA RADI POBOLJŠANJA DOSTUPNOSTI

Zbog mogućnosti objavljivanja digitalnih preslika putem mreže, digitalizacija u potpunosti mijenja koncept pristupa i korištenja kulturno-povijesne baštine. Dostupnost građe na daljinu demokratizira pristup gradi i podupire ciljeve obrazovanja i znanstveno-istraživačkog rada. Digitalizacija omogućuje bolje korištenje građe u interdisciplinarnim i multidisciplinarnim znanstvenim i obrazovnim projektima koji istražuje određenu temu, autora, povijesno razdoblje itd.. Digitalne preslike mogu se koristiti u predstavljanju građe i ustanove putem interneta. Ustanove koje na taj način nude svoje sadržaje mogu računati na to da će se proširiti njihova korisnička zajednica, da će njihova ponuda postati vidljivija i privući nove vrste korisnika.³³

Upravo iz razloga poboljšanja dostupnosti smo krenuli u ovaj mini projekt i smatramo kako nema opravdanja za ne pokretanjem masovne digitalizacije građe koja bi se mogla smjestiti u okvir građe za koju je od velike važnosti poboljšanje dostupnosti i jednostavnosti uporabe poput ove bibliografije, naravno tu je vječito moderni problem financiranja koji se opet može riješiti naplaćivanjem simboličnih naknada za pristup digitaliziranim sadržajima, slično fotokopiranju u knjižnicama.

DIGITALIZACIJA RADI STVARANJA NOVIH PROIZVODA I USLUGA

Digitalizacija omogućuje ponudu novih usluga korisnicima koje ne bi bile moguće ili bi bile teško izvedive izvan električnog okruženja, a na-

³¹ Ministarstvo kulture Republike Hrvatske, *Smjernice za odabir građe za digitalizaciju*, studeni 2007. URL: http://www.kultura.hr/hr/content/download/596/7925/file/smjernice_odabir.pdf (1.9.2010)

³² isto.

³³ isto.

staju objedinjavanjem digitalnog sadržaja, podataka o digitalnoj građi u stvarnim ili virtualnim sustavima. Digitalne preslike mogu se višestruko koristiti u okviru različitih službi i za različite usluge pa se pri odabiru za digitalizaciju može uspostaviti opće pravilo da građa koja omogućuje više primjena ima veći prioritet. Jedan od proizvoda sustavnog bavljenja digitalizacijom bit će i razvoj infrastrukture, znanja i iskustva, koji također mogu poslužiti uvođenju određenih specijaliziranih usluga.³⁴

Mogućnosti razvoja usluga baziranih na digitalnim sadržajima su toliko velike da bi cijela jedna generacija studenata imala posla na razvoju i planiranju interaktivnih aplikacija i usluga samo jedne knjižnice. Zamislimo samo npr. cijelu zbirku novina pohranjenu u Muzeju Slavonije pretraživu po uzoru na recimo Google-ovu tražilicu, ili neku od tisuća besplatnih skripti za pretraživanje dostupnih na internetu.

DIGITALIZACIJA RADI UPOTPUNJAVANJA FONDA I SURADNJE

Nabava digitalnih preslika građe iz drugih ustanova ili privatnih zbirki jedan je od načina upotpunjavanja fondova ustanova ili izgradnje zbirki za koje se procjenjuju da su potrebne s obzirom na korisnike, svrhu ili područje djelovanja ustanove. U takvim će slučajevima digitalna preslika biti i jedini oblik (osim možda kopija ispisanih i na pisaču) u kojem građa postoji u ustanovi. Stoga se postavljaju isti zahtjevi u svezi sa zaštitom i dostupnošću, uz poštivanje autorskog prava i drugih prava vlasnika izvornika.³⁵

Ponovno gledajući s finansijskog aspekta koji je danas tako zgodno postao isprika i alibi za sve i svašta, pitam se koliko bi ta ista digitalizacija bila skupa kada bi jedna knjižnica koja digitalizira npr. rijetku jedinicu građe uz simboličnu naknadu ustupila digitaliziranu jedinicu pet drugih knjižnica u hrvatskoj i time isplatiла projekt a upotpunila fond drugih knjižnica. Valja spomenuti kako bi i knjižnicama koje upotpunjaju fond bila jeftinija nabavka građe na ovaj, digitalni način (nakon inicijalnih ulaganja koja zbilja nisu velika).

DIGITALIZACIJA NA ZAHTJEV

Premda se iza svih navedenih razloga digitalizacije, a osobito dostupnosti i demokratizacije pristupa građi, jasno vide korisnici, potrebno je posebno istaknuti digitalizaciju na zahtjev korisnika od kojih, ponajprije u knjižnicama i arhivima, dolazi velik dio zahtjeva za digitalizacijom. Ako se

³⁴ Ministarstvo kulture Republike Hrvatske, Smjernice za odabir građe za digitalizaciju, studeni 2007. URL: http://www.kultura.hr/hr/content/download/596/7925/file/smjernice_odabir.pdf (1.9.2010)

³⁵ Ministarstvo kulture Republike Hrvatske, Smjernice za odabir građe za digitalizaciju, studeni 2007. URL: http://www.kultura.hr/hr/content/download/596/7925/file/smjernice_odabir.pdf (1.9.2010)

uspostavi zadovoljavajuća infrastruktura i primjerene kompetencije, sva-ka bi ustanova trebala biti u mogućnosti ispuniti zahtjev za digitalizacijom građe. Svakako, digitalizacija na zahtjev ne smije biti jedini način određivanja odabira jer skup na taj način izrađenih digitalnih preslika ne osigurava izgradnju osmišljene digitalne zbirke. U izradi popisa prioriteta, ustanove moraju naći mjeru da zadovolje planirane - redovne programe digitalizaci-je i ispune korisničke zahtjeve.³⁶

Dovoljno je reći kako smo kao studenti često došli u situaciju da neke di-jelove literature sami prenositi u digitalni oblik, prednosti digitalnog zapi-sa su višestruke i već su u više navrata spomenute u prijašnjem dijelu rada.

METODA DIGITALIZACIJE I ZAVRŠNI REZULTAT

Digitalizacija ovog djela, kao što sam već spomenuo, nije za primarni cilj imala prezervaciju samog djela od propadanja jer se ne radi o posebno sta-joj jedinici građe niti o posebno rijetkoj knjizi već je misao vodilja bila učiti ovu bibliografiju jednostavnijom za korištenje; uštedjeti vrijeme i učiniti ju dostupnijom korisnicima. Dakle pošto je knjiga u vrlo dobrom stanju nije bilo potrebno koristiti se skupim³⁷ i nedostupnim³⁸ skenerima prilagode-nim staroj i osjetljivoj građi, već sam se koristio nešto kvalitetnijim flatbed skenerom.

Za samo skeniranje ove knjige korišten je fotokopirni stroj Canon iR8070 te njegova nativna aplikacija za skeniranje. Uz ručno unošenje postavki od kvalitete skena na 600dpi te 8-bitnu gray-scale kombinaciju boja (najpo-godnije za OCR) nakon nekoliko testnih stranica jasno je bilo kako je rezolucija 1303*885 pixela i JPEG format dobra kombinacija koja zauzima malo mesta na disku a OCR program koji sam koristio je davao jednake rezultate pri ovoj rezoluciji kao i na svim ostalim mnogo većima koje zau-zimaju mnogo više diskovnog prostora i usporavaju proces. Jednako tako, program koji sam koristio se potpuno jednako ponašao pri korištenju JPEG i TIFF datoteka, pa je logičan izbor bilo korištenje JPEG-a koji zauzimaju višestruko manje diskovnog prostora. Pošto sam se koristio kvalitetnim skenerom A3 formata a dimenzije knjige su to dopuštale, skenirali smo dvi-je stranice od jednom te dobili 293 JPEG datoteka koje sadrže pun opseg od 587 stranica ove bibliografije. Prosječna veličina svake datoteke je 484 KB što po stranici znači iznosi 242 KB. Uz nešto drugačije postavke ta se veličina mogla svesti i na više nego dvostruko manje ali pošto sveukupna veličina od 142 MB u današnjim pojmovima zbilja nije mnogo i imajući na

³⁶ isto

³⁷ ImageWare – Complete solutions for digitisation projects. URL: <http://www.imageware.de/en/systems/> (1.9.2010)

³⁸ Coyle, Karen. Mass digitization of books. URL: <http://www.kcoyle.net/jal-32-6.html> (1.9.2010)

umu da smo digitalizirao samo jednu knjigu a ne stotine, uzeli smo ovaj omjer veličine datoteke i kvalitete digitalnog zapisa kao zadovoljavajući. Uvezši u obzir da je namjena ovog projekta poboljšati pretraživost a ne dati vjernu presliku izvornika, sama kvaliteta skena je sekundarna a primarni cilj je dobiti što bolje OCR rezultate koji će se i koristit.

Nakon dobivenih JPEG datoteka krenuo sam na provođenje OCR-a. Program koji sam koristio za ovaj postupak je „ABBYY FineReader 9.0 Professional Edition“³⁹. Glavni razlog za odabir ovoga programa je što je lokализiran i sposoban za rad sa raznim svjetskim jezicima što je bilo od velike važnosti jer ih u Osječkoj bibliografiji ima pet. U međuvremenu se pojavila i nova inačica ovog programa „ABBYY FineReader 10.0“ koja pruža nove pogodnosti, te unaprjeđuje već postojeće. Nova inačica ima podršku za nove operativne sustave, višejezgrene procesore i nova programska rješenja koja se u međuvremenu pojavila na tržištu.

Nakon završenog skeniranja i OCR obrade digitalizirana jedinica ustupljena je Muzeju Slavonije u Osijeku na daljnje raspolaganje.

³⁹ OCR software for text recognition. URL: <http://finereader.abby.com/> (1.9.2010)