

DIGITALIZACIJA GRAĐE

HRVOJE STANČIĆ

Filozofski fakultet, Zagreb

Digitalizacija je danas jedna od vrućih tema kako u arhivima, tako i u muzejima i knjižnicama. Stoga ću u ovom izlaganju govoriti o samom procesu digitalizacije i o problemima koji se pritom javljaju.

Na početku treba istaknuti da je u ovom izlaganju riječ o digitalnim arhivima, digitalnim muzejima i digitalnim knjižnicama, ali je upravo zadnji termin u medijima nekako najzastupljeniji. Digitalna knjižnica je, prema Watersu, "organizacija koja osigurava sredstva i osoblje za odabir, strukturiranje, internacionalni pristup, interpretiranje, distribuiranje, očuvanje integriteta i osiguranje postojanosti kroz vrijeme digitalnih kolekcija kako bi se osiguralo brzo i ekonomično pristupanje, pregledavanje i korištenje".¹ Smatram da se tako mogu definirati i termini digitalni arhiv i digitalni muzej.

Također je potrebno istaknuti da se digitalizacijom građe pojam očuvanja počinje dijeliti na dva dijela: očuvanje informacijskog sadržaja, tj. informacije koju određeni dokument nosi, te očuvanje fizičkog objekta kao nositelja informacije. Informacijski se sadržaj, dakle, digitalizira i sprema odvojeno od objekta-nositelja.

Četiri su prednosti digitalizacije:

- Digitalizacijom se lakše može očuvati od oštećenja rijetka i osjetljiva građa, a istodobno osigurati pristup informacijskom sadržaju koji ona prenosi.
- Druga je prednost jednostavnost pristupa. Kada je dokument jednom digitaliziran i stavljen na mrežu, više ga osoba istodobno može čitati iz bilo kojeg dijela svijeta.
- Treća prednost građe u digitalnom obliku jest ta da ona zauzima milimetre na disku umjesto metre na policama.
- Četvrta je prednost mogućnost neograničena kopiranja, a da se pritom ne gubi kvaliteta.

1. PROCES DIGITALIZACIJE

Koncepcijski problemi i odabir prave tehnologije za digitalizaciju vrlo su kompleksni. Svaki stupanj u procesu digitalizacije zahtijeva zasebna koncepcijska i hardversko-softverska rješenja

koja izravno ovise o građi koja se obrađuje. Proces se sastoji od ovih koraka: 1. odabir dokumenata, 2. unos, 3. obrada, 4. zaštita, 5. pohrana i prijenos, te 6. pregled i korištenje. Krenimo redom.

1.1 ODABIR DOKUMENATA

Na početku procesa digitalizacije potrebno je odlučiti koju građu digitalizirati. Kada kažem *koju građu*, mislim *koju građu najprije* digitalizirati jer se danas sve može digitalizirati uz više ili manje napora i veće ili manje troškove. Odabir, dakako, ovisi prije svega o vrsti ustanove, njezinu programu i ciljevima koje želi digitalizacijom postići. Što se tehničke strane tiče, najprije treba provjeriti za koju građu ne postoje autorska prava. Takva se građa smije digitalizirati bez plaćanja naknade za digitalizaciju, omogućavanje pristupa itd. Zatim, najbolje je odabrati veće količine materijala koji je jednostavniji za digitalizaciju i za koji se očekuje manje problema pri obradi. Tako je, na primjer, bolje odabrati knjige nego stare karte koje sadrže sliku, tekst i možda pokoju noticu rukom zabilježenu na margini već požutjela papira. Digitalizirati najprije najpopularnija djela dobar je način za stjecanje šire popularnosti i potpore. U američkoj je Kongresnoj knjižnici, na primjer, roman *Prohujalo s vihorom* tražen u digitalnoj verziji više od bilo kojega teksta o Prvome svjetskom ratu.

1.2 UNOS PODATAKA

Fizičke dimenzije i osjetljivost građe koja se želi skenirati, uvjetuju odabir opreme.

Pri skeniranju tekstova dokumenti koji su na zasebnim listovima papira standardne veličine mogu se uglavnom automatski skenirati koristeći stolne skenere i uvlakače papira. Uvezane sveske treba skenirati odozgo kako bi se spriječilo oštećenje.

Za građu većih formata mora se koristiti visokokvalitetna digitalna kamera kojom se predložak također digitalizira okrenut licem prema gore. Rasvjeta prilikom digitalizacije kamerom vrlo je bitna, pogotovo kada se digitaliziraju predlošci pri čijem se osvjetljavanju javljaju zrcalni odbljesci. Tako je, na primjer, kod različitih srebrnih površina potrebno istodobno osvjetljavanje sa više strana.

Tekst se može pretvarati u digitalni oblik prijepisom ili skeniranjem. Prijepis je dugotrajan, iscrpljujući i vrlo skup način prijenosa teksta u digitalni oblik. Druga je mogućnost skeniranje originala ili već postojećeg mikrofila. Međutim, u tom slučaju kao rezultat dobiva se slika teksta, a ne tekst koji se može obrađivati. Tada je potrebna dodatna obrada programom za optičko prepoznavanje slova, koji sliku teksta pretvara u obradivi tekst.

Pri digitalizaciji slikovne građe treba odrediti najveću dopuštenu izloženost građe svjetlu, te paziti pri odabiru skenera i osvjetljenja za kameru da njihovi svijetleći elementi ne proizvode ultravioletnu komponentu svjetla, jer je ona najzaslužnija za većinu blijedenja i degradacije

umjetničkih slika. Budući da se digitalizacijom želi sačuvati informacijski sadržaj predloška, digitalna slika mora biti kvalitetna. To znači da ona mora biti skenirana u visokoj razlučivosti (najmanje 600 dpi u 24-bitnoj boji za manje formate, te 300 dpi za veće formate). Što je slika skenirana u više razlučivosti, to zauzima više mjesta na disku i neprikladnija je za prijenos mrežom. Primjera radi, slika veličine A4 formata papira, skenirana u razlučivosti od 600 dpi koristeći 8 bitova (256 nijansi) za svaku od tri osnovne boje (znači 24-bitna boja), zauzima diskovni prostor veličine 96 Mb. Preporuča se izrada dviju verzija svake digitalizirane slike - jedna u boji, visoke razlučivosti koja mora biti što vjernija originalu i na kojoj se nalazi samo ilustracija, te druga u skali sivih tonova, niže razlučivosti koja obuhvaća cijelu sliku zajedno s otisnutim ili rukom dodanim noticama. Obje se slike poslije, radi pogodnosti za prijenos mrežom, dodatno obrađuju, dok se originali čuvaju u izvorno digitaliziranom obliku.

1.3 OBRADA PODATAKA

Budući da je svaka stranica teksta nakon skeniranja pohranjena kao slika, tekst se ne može obrađivati, pretraživati ni indeksirati. Potrebno je, dakle, sliku pretvoriti u tekst. To se postiže takozvanim OCR programima, tj. programima za optičko prepoznavanje slova. Prepoznavanje teksta može se obavljati pri samom skeniranju, ali i naknadno. OCR programi razlikuju bijele od crnih površina teksta, te stvaraju tekstualnu datoteku. Problemi nastaju kod nedovoljno kontrastnih predložaka, tekstova s čestim tipografskim promjenama (na primjer u rječnicima i enciklopedijama), tekstova koji sadrže znakove svojstvene drugim jezicima, tekstova sa zastarjelim ili staromodnim oblicima pisma. Tako je, na primjer, povijesne dokumente i stare novine, da ne govorimo o rukom pisanim dokumentima, gotovo nemoguće brzo, kvalitetno i ekonomski isplativo skenirati. Želi li ih se digitalizirati, takve dokumente treba prepisati.

Današnji OCR programi kvalitetne predloške iz slike u tekst pretvaraju s prosječnom točnošću od 99,95 posto. To još uvijek znači da ručno treba ispraviti jednu do dvije pogreške po stranici. Treba spomenuti da se, ako se točnost želi podići na 99,99 posto, troškovi udvostručuju.²

Nakon uspješna optičkog prepoznavanja slova, dobili smo tekst koji tek trebamo urediti tako da se u svakom trenutku može znati na kojoj se stranici u tiskanom izdanju nalazi dio koji čitamo. Mnoge knjige imaju ilustracije, slike, grafove, tablice itd. koje također treba digitalizirati i staviti ih na njihovo mjesto u digitalnoj verziji knjige. Ne zaboravimo, tu su i podnožne bilješke.

Troškovi su digitalizacije knjige od tri stotine stranica skeniranjem oko \$30, a prijepisom oko \$600. Pretvaranje pak teksta u hipertekstualni format stoji oko \$900, dakle trideset puta skuplje od običnog skeniranja!

Slikovni se zapisi, zbog velikih memorijskih zahtjeva, moraju komprimirati. U takvu su obliku pogodniji za pohranu i prijenos. Dvije su vrste kompresije: bez gubitaka i s gubicima. Kompresija

bez gubitaka čuva potpunu slikovnu informaciju. Kod kompresije s gubicima (najpoznatiji je JPEG format) možemo odabrati željeni stupanj kompresije. Što je stupanj kompresije viši, rezultirajuća je datoteka manja, ali je manja i kvaliteta slike, i obrnuto. Dakle, treba odrediti optimalni odnos kompresija-veličina-kvaliteta.

1.4 ZAŠTITA PODATAKA

Kako korisnik saznaje da je sadržaj koji vidi na zaslonu autentičan? Upravo zato je u digitaliziranu građu potrebno ugraditi zaštitne oznake poput onih koje se primjenjuju pri zaštiti novčanica. Mogu se koristiti digitalni vodeni žigovi, sustavi kriptiranja i druge tehnike dokaza autentičnosti. U njihovu dublju analizu neću ulaziti jer su one zasebna tema izlaganja, no ipak je bitno spomenuti da je provjera autentičnosti veoma brz proces.

1.5 POHRANA I PRIJENOS PODATAKA

Još pri odabiru programa za obradu digitaliziranih podataka treba paziti da taj program podržava potreban format zapisa, jer svaki format ne pruža one opcije koje su potrebne pri pohrani digitalizirane građe. Naime, u datoteci digitalizirane slike, osim sama zapisa, moraju biti zapisane i kontrolne informacije. Zbog toga se preporučuje koristiti TIFF format zapisa jer taj format podržava komprimirani i nekomprimirani zapis, posjeduje zaglavlje u koje se zapisuju informacije o slici čime se ubrzava pretraživanje i osigurava mogućnost kasnije analize, te je *de facto* standard.

Pohrana i osiguravanje dostupnosti digitalne građe u velikim sustavima zaseban je problem. Budući da se građa pristupa preko mreže, razvili su se takozvani *on-line*, *near-line* i *off-line* sustavi. Kao primjer mogu se navesti pojedinačni hard-diskovi, polja hard-diskova, te kombinacije hard-diskova, DVD-ova, CD-ROM-ova i magnetskih vrpca, koji se upravljaju takozvanim hijerarhijskim sustavom za upravljanje datotekama koji može čak upravljati i robotskim sustavom za manipulaciju nosačima podataka. Taj se sustav naziva HMS sustav (Hierarchical File Management System), te se jednostavno može opisati kao "virtualni disk". On u sebi ujedinjuje *on-line*, *near-line* i *off-line* koncept.

Treba spomenuti da cijena spremanja 300 stranica u digitalnom obliku na hard-disk iznosi oko \$2.

1.6 PREGLED I KORIŠTENJE

Problemi pregleda i korištenja digitalizirane građe ovise o tome hoće li se građa samo pregledavati na zaslonu ili će se i ispisivati. Ako se radi o tekstualnom tipu građe, većina današnjih pisaa ima kvalitetan crno-bijeli ispis. No ako se radi o slikovnoj građi, posebnu pažnju treba obratiti na odabir pisaa jer on mora davati zadovoljavajuće rezultate, te

eventualno imati mogućnost ispisa većih formata. Ako se radi o ispisu slika u nijansama sive boje, razlučivost ispisa mora biti jednaka ili veća od razlučivosti slikovnog zapisa, premda ispisi pri razlučivosti od 600 dpi daju tek kvalitetu usporedivu s novinskim slikama. Za ispis u boji, koji je zasebna problematika, kvalitetne rezultate daju pisači koji rade na principu sublimacije.³

2. ODRŽAVANJE DIGITALIZIRANE GRAĐE

Brz razvoj tehnologije ujedno znači i brzo zastarijevanje njezinih prethodnih oblika. Tko danas može pročitati zapis bušenih kartica? Rijetki posjeduju disketni pogon veličine 5,25". Tko zna koliko ćemo dugo koristiti 3,5" diskete? Pojavom DVD-a, i CD-ROM-ovi dolaze u pitanje. Ovime želim istaknuti da se u projekt digitalizacije građe moraju ugraditi i tehnike za uspješnu migraciju podataka na nove medije. Troškovi su migracije visoki, a proces ponekad dugotrajan. Naime, migrirati podatke na novi medij ne znači samo fizičko kopiranje. Ponekad treba konvertirati podatke u drugi zapis, a sam proces treba prethodno dobro testirati kako bi se sačuvala kvaliteta. Nakon migracije treba nasumce odabrati određeni postotak zapisa, te provjeriti odgovara li njihova kvaliteta originalnim zapisima.

Što se tiče postojanosti medija, mikrofilm je u velikoj prednosti jer ispravno obrađen i pohranjen veoma dugo traje. Stoga se nameće pitanje da li digitaliziranu građu ujedno snimati i na mikrofilm. Ako da, da li najprije digitalizirati, pa digitalnu verziju računalno snimiti na mikrofilm, ili najprije snimiti mikrofilm, pa digitalizirati s mikrofilma. Svaki pristup ima svoje prednosti i nedostatke, pa je u svakom konkretnom slučaju potrebno provesti prethodno testiranje.

O digitalizaciji treba razmišljati u relativnim, a ne u apsolutnim okvirima, tj. kao o ciklusima pohrane podataka, jer vrijeme upotrebljivosti jednog medija zapravo je vrijeme do novog ciklusa migracije podataka.

Naposljetku, skrenuo bih pažnju i na ubrzan razvoj HD-Rosette i HD-ROMa. To su novi mediji visoke gustoće zapisa, kapaciteta do 165 Gb po jednoj strani diska veličine 2". Zapis je na disku nemagnetski, pa je stoga dugotrajan (i do tisuću godina) i otporan na vanjske elektromagnetske utjecaje. Podržan je analogni i digitalni zapis na jednom te istom disku. Tako se uz pomoć optičkog mikroskopa visoke rezolucije mogu iščitavati analogni zapisi. Ta je tehnologija u razvoju usporediva s mikrofilmom po svojoj neovisnosti o neprestanim novim verzijama softvera, te je treba uzeti kao ozbiljna kandidata za spremanje građe i zbog njezinih prednosti što se tiče dugotrajnosti medija, prenosivosti i visoke gustoće zapisa.⁴

3. ZAKLJUČAK

Digitalizacijom građe bitno se olakšava pristup, pregled i pretraživanje građe. Proces digitalizacije vrlo je složen pa veliku pozornost treba obratiti odabiru tehnika i tehnologija koje će se pritom koristiti, te proračunu potrebnog vremena i sredstava. Poželjno je krenuti s digitalizacijom jednostavnije građe kako bi se izbjegli problemi na samom početku. Preporučljivo je digitalizaciju prepustiti profesionalcima izvan matične organizacije ili zatražiti njihov stručni savjet. Treba razvijati suradnju s drugim ustanovama koje su već krenule u digitalizaciju ili bar provode početna istraživanja i testiranja, razmjenjivati znanja i iskustva, te izbjeći digitalizaciju onoga što je već digitalizirano. Budući da je danas produkcija knjiga nezamisliva bez uporabe računala, treba također ispitati uvjete i mogućnosti pribavljanja novih izdanja u digitalnom obliku od samih izdavača, kao nositelja autorskih prava. Uza sve to, potrebno je stručno usavršavati vlastite kadrove koji bi opća znanja o procesu digitalizacije trebali stjecati već tijekom svog obvezatnog studija.

BILJEŠKE

1. Cleveland, Gary. *Digital libraries: definitions, issues and challenges*, ožujak 1998., str. 2 digitalne verzije, <http://www.ifla.org/ifa/VI/5/op/udtop8/udtop8.htm>
2. Jann, Lynn-George. *Digitization: a literature review and summary of technical processes, applications and issues*, svibanj 1996., str. 12 digitalne verzije, http://www.library.ualberta.ca/library_html/libraries/law/digit1.html
3. Sublimacija je proces prelaženja čvrste tvari zagrijavanjem direktno u plinovitu i ponovno kondenziranje para u čvrstu tvar.
4. Zahvaljujem prof. dr. Miroslavu Kržaku što mi je skrenuo pažnju na ovu tehnologiju.

LITERATURA

- Arms, William Y. *Key concepts in the architecture of the digital library*. // D-Lib magazine, srpanj 1995. <http://www.dlib.org/dlib/july95/07arms.html>
- Besser, Howard; Jennifer Trant. *Introduction to imaging: issues in constructing an image database*. The Getty Information Institute Imaging Initiative, 1995. http://www.gii.getty.edu/intro_imaging/Tbl.html
- Cleveland, Gary. *Digital libraries: Definitions, issues and challenges*, ožujak 1998. <http://www.ifla.org/ifa/VI/5/op/udtop8/udtop8.htm>
- Cleveland, Gary. *Overview of document management technology*, lipanj 1995. <http://www.ifla.org/ifa/VI/5/op/udtop2/udtop2.htm>
- D'Amato, Donald P.; Rex C. Klopfenstein. *Requirements and options for the digitization of the illustration collections of the national museum of natural history*, ožujak 1996. <http://www.nmnh.si.edu/cris/techrpts/imagopts/contents.html>
- Dollar, Charles M. *Archival theory and information technologies. The impact of information technologies on archival principles and methods*. University of Macerata, 1992.
- Graham, Peter S. *Intellectual preservation: electronic preservation of the third kind*, ožujak 1994. <http://palimpsest.stanford.edu/byauth/graham/intpres/#fn9>
- Hunter, Philip. *Text encoding and the digitization of text*, lipanj 1995.

<http://www.presence.co.uk/hippeis/ELA/digitb.htm>

IFLA *position paper on copyright in the electronic environment*, listopad 1996. <http://www.ifla.org/ifla/V/ebpb/copy.htm>

Jann, Lynn-George. *Digitization: a literature review and summary of technical processes, applications and issues*, svibanj 1996. http://www.library.ualberta.ca/library_html/libraries/law/digit1.html

Kenney, Anne R. *Digital to microfilm conversion: a demonstration project 1994-1996*.
<http://www.library.cornell.edu/preservation/com/comfin.html>

Kuny, Terry. *An introduction to digitization technologies and issues*, listopad 1995.
<http://www.nlc-bnc.ca/publications/netnotes/notes14.htm>

Lesk, Michael. *Going digital*. // Scientific American 3, 1997. <http://www.sciam.com/0397issue/0397lesk.html>

Norsam Technologies Inc. *Introducing a revolution in ultra high density data storage*, 1997.
<http://www.norsam.com/>

Siemens. *Networking and digitization. Archiving and managing digital information*, rujan 1997.
<http://www.svb.nl/project/harmonica/Deliverables/D33.htm>

Sutton, Clare; Robina Clayphan. *Transmission of data*, prosinac 1996. <http://hosted.ukoln.ac.uk/biblink/wp5/>

Waters, Donald J. *Electronic technologies and preservation*, lipanj 1992.
<http://www.clir.org/pubs/reports/waters/waters2.html>

Weber, Hartmut; Marianne Dörr. *Digitization as a method of preservation*. Amsterdam: European Commission on Preservation and Access, srpanj 1997.

Zillhardt, Sonia. *Bibliotheca Universalis: a G-7 global information pilot project*, kolovoz 1998.
<http://www.ifla.org/IV/ifla64/031-98e.htm>