Prikaz ustroja sustava za digitalizaciju mikrofilmova Nacionalne knjižnice Republike Češke, Prag, 12–16. svibnja 2003.

1. Općenito o Nacionalnoj knjižnici Republike Češke i Odjelu za zaštitu knjižnoga fonda

Zadatak Nacionalne knjižnice Republike Češke (u daljnjem tekstu: Knjižnica) je skupljanje, zaštita i osiguravanje dostupnosti knjižnoga gradiva, objavljenoga u Republici Češkoj, literature na češkom jeziku objavljene u inozemstvu, kao i odabranih stranih publikacija. Ukupan knjižni fond Knjižnice procjenjuje se na 6 milijuna svezaka. Knjižnica se od 1996., otkad je završeno preuređenje zgrade u Hostivaru, nalazi na dvjema lokacijama. U Hostivaru se nalazi sjedište Odjela za zaštitu knjižnoga fonda, unutar kojega djeluju Odsjek za mikrografiju, Odsjek za digitalizaciju, Odsjek za restauraciju i Odsjek skrbi za moderne publikacije.

2. Mikrofilmiranje u Nacionalnoj knjižnici Republike Češke

Mikrofilmiranje je u Knjižnici započelo još krajem 1940-ih zahvaljujući poklonu Rockfellerove zaklade, kad je Knjižnici darovana stara kamera Rekordak. Tada se pristupilo mikrofilmiranju najvažnijih rukopisa. Mikrofilmiranje novina i časopisa započelo je krajem 1960-ih. Zbog relativno niskoga kapaciteta mikrofilmiranja ondašnjih kamera, do druge polovice 1980-ih mikrofilmirano je svega oko 1500 svezaka rukopisa i nekoliko važnijih novinskih naslova (Rudé právo, Právo lidu, Bohemie itd.). Mikrofilmovi izrađeni u to vrijeme nisu slijedili ISO standarde, te su snimani bez test-snimaka i podataka o stupnju umanjenja, što danas uvelike otežava postupak njihove digitalizacije.

Za mikrofilmiranje su u početku rabljeni filmovi ORWO DK 5, poslije ORWO MA 8, a za korisničke kopije filmovi FOMA Kinopozitiv; u početku perforiran, potom neperforiran. Snimanje se obavljalo na kamerama DOKUMATOR DA 4, poslije na tipu DOKUMATOR DA 5, starijem tipu kamere RECORDAK i novijoj kameri RECORDAK MRD 2. Razvijanje filmova obavljalo se ručno, a zatim uređajima MEOLAB.

Godine 1983. Knjižnica počinje snimati i mikrofiševe na uređaju Pentakta. Tako je snimljeno oko 70 naslova časopisa i 300 starih tiskanih knjiga. Od početka 1990-ih prestalo se snimati na mikrofiševe, jer se film ORWO prestao proizvoditi, a KODAK-ovi filmovi i ostale tada poznate marke filmova nisu podnosile razvijanje u uređaju PENTAKTA E 120.

Novi uređaji u laboratoriju Knjižnice nabavljeni su zahvaljujući projektu CASLIN[1] i američkoj Zakladi A. W. Mellona. Veći dio Odjela za mikrografiju preselio se 1996. u novouređenu zgradu Središnjega repozitorija u Hostivaru. Knjižnica trenutačno posjeduje kamere tipa Zeutschel, Gratek, Elka, tri kamere DA 5 i jednu DA 7.

---

[1] CASLIN – fokus ovoga projekta bio je uspostava nacionalne mreže knjižnica. Ta mreža trebala bi osigurati domaćim i stanim korisnicima lagan i brz pristup informacijama u češkim i slovačkim knjižnicama. Rezultat je projekta uspostava središnjega nacionalnoga kataloga čeških knjižnica.

Uz kamere, tu su i uređaji za razvijanje, kopiranje i druga suvremena dijagnostička i kontrolna pomagala. Za skeniranje mikrofilmova Knjižnica je nabavila dva skenera: SunRise, kupljen 1998. godine, i Wicks & Wilson 4100, kupljen krajem 2001. godine.

2.1. Postupak mikrofilmiranja

Za mikrofilmiranje se odabiru dokumenti koji su najviše oštećeni, bilo zbog kemijske degradacije papira, bilo zbog prečeste uporabe. Uglavnom se mikrofilmiraju novine i časopisi, rukopisi i ostali slični dokumenti. Dio kapaciteta posvećen je mikrofilmiranju regionalne i specijalne periodike koju posjeduju druge institucije (knjižnice, arhivi i muzeji).

Gradivo se (najčešće uvezani svesci časopisa i novina) prije mikrofilmiranja pomno priprema. Svaki se svezak prelistava i podrobno pregledava ne bi li se uočile tipične pogreške, kao što su pogrešna paginacija, pogrešna datacija, pogrešan uvez. Provjeravaju se podatci o naslovu i podnaslovu, broju i godini izdanja itd. Svi podaci prikupljeni tijekom ove kontrole upisuju se u točno određeni obrazac. Poslije se iz tih podataka formiraju metapodatci u formatu DOBM SGML (o metapodatcima detaljnije u odjeljku 4). Provjeru simultano obavljaju dvije osobe, kako bi se u još većoj mjeri izbjegli previdi neke pogreške. Ovisno o debljini sveska, za kontrolu im je potrebno od jednog do tri sata.

Poslije pripreme stvara se detaljni bibliografski opis, koji se zajedno s identifikacijskim i sadržajnim predlošcima snima na početak filma. Uputnice o pogrešnoj paginaciji, uvezu i drugim pogreškama označuju se dvojezično – na češkom i engleskom jeziku. Uputnice se snimaju na mjestu gdje se dotična pogreška na originalu pojavljuje. Sve spomenute uputnice i predlošci (identifikacijski i sadržajni) otisnuti su tako da se mogu pročitati s mikrofilma i bez pomoći mikročitača ili povećala.

Posebna se pažnja posvećuje i samom procesu mikrofilmiranja i njegove kemijske obrade, naročito zbog činjenice da je u čitav proces uvedena i faza digitalizacije, zbog koje zahtjevi za kvalitetno izrađenim mikrofilmom rastu. Čitav svitak mikrofilma morao bi imati ravnomjernu gustoću pozadine (D^{max} – *density maximum*), što je, s obzirom na velike razlike u obojenosti papira, nemoguće postići.

Kontrola kakvoće mikrofilma obavlja se na temelju test-obrazaca ISO 1 i ISO 2, koji se snimaju na početku mikrofilma. Testni se obrasci pregledavaju mikroskopom s povećanjem 100 puta.

Nakon mikrofilmiranja i izrade posrednoga negativa i korisničke kopije, slijedi digitalizacija.

Važno je spomenuti da je Knjižnica član EROMM-a (*European Register of Microform Masters*). EROMM je međunarodna baza podataka koja sadržava podatke o mikrofilmiranoj knjižničnoj građi. Baza je dostupna online, sa svrhom da se biblioteke i ostale korisničke obavijesti o tome što je od građe već mikrofilmirano, kako bi se spriječilo dupliciranje posla, tj. ponovno mikrofilmiranje iste građe, čime se štedi na vremenu i novcu. Da bi neko mikrofilmirano gradivo bilo uneseno u bazu podataka EROMM-a, mikrooblik mora biti izrađen i pohranjen prema točno određenim tehničkim

standardima.[2] Sve donedavno, u EROMM-u su postojali podatci samo za mikrooblike, no oni se postupno proširuju i na digitalne kopije.

3. Digitalizacija u Nacionalnoj knjižnici Republike Češke

Knjižnica provodi digitalizaciju svojih fondova na dva načina. Prvi način je tzv. «izravna» digitalizacija. Tim se načinom digitaliziraju rukopisi i rijetki spisi. Svrha je takve digitalizacije postići visokokvalitetne snimke u boji kako bi se dobila što bolja kopija rukopisa, koji su često ilustrirani. Program izravne digitalizacije izniknuo je kao posljedica sudjelovanja Knjižnice u projektu «Memory of the World». Knjižnica ga provodi u suradnji s tvrtkom Albertina Icome Praha.

Drugi način digitalizacije jest digitalizacija mikrofilma. Program je nastao zahvaljujući projektu «Digitalizacija mikrooblika», a koji je postojao 1997–1999. Spajanjem mikrofilma i digitalnih zapisa ostvaruje se tzv. hibridna tehnologija, koja rabi prednosti obiju tehnologija. Mikrofilm svojom dokazanom postojanošću osigurava trajnu pohranu sadržaja dokumenta, a digitalni zapis korisnicima omogućuje lakši pristup sadržaju.

Proces digitalizacije osim skeniranja mikrofilmova uključuje i tvorbu metapodataka, spajanje metapodataka s nastalim skenovima i omogućivanje pristupa digitalnim zapisima preko Interneta ili CD-R medija.

3.1 Priprema dokumenata i izrada metapodataka

Podatci s obrazaca koji se ispunjavaju prilikom pripreme dokumenata za mikrofilmiranje (a u koji su uneseni podatci o strukturi predloška i eventualnim nepravilnostima unutar njega) prepisuju se u računalo kao neformatirani tekst, koji se zatim automatski konvertira u DOBM SGML oblik. Time je stvoren metapodatkovni opis strukture podataka. Upotreba metapodataka igra ključnu ulogu u dugoročnom očuvanju pristupa digitalnim dokumentima.

3.2. Skeniranje mikrofilmova i obrađivanje skenova

Za skeniranje mikrofilmova Knjižnica upotrebljava skener SunRise, koji može skenirati mikrofilmske svitke širine 16 i 35 mm, mikrofiševe u razlučivosti do 600 DPI i skener Wicks and Wilson 4100, koji može skenirati mikrofilmske svitke širine 16 i 35 mm u razlučivosti 200 i 400 DPI. Nakon skeniranja, skenovi se učitavaju u grafički preglednik ACDSee, gdje se vizualno provjerava njihova ispravnost i kakvoća. Po potrebi se skenovi dodatno obrađuju, poravnavaju te se uklanjaju suvišni rubovi slike. Na samom početku digitaliziranja, Knjižnica je proizvodila slike u jednobitnom TIFF formatu, no poslije se prešlo na isključivo skeniranje u 8-bitnoj sivoj skali u formatu JPEG. Kako Knjižnica proizvodi slike u razlučivosti do 600 DPI, prosječna slika u JPEG formatu može biti velika i do nekoliko MB. Pri toj veličini slike postaju nepogodne za distribuciju preko Interneta. Problem je riješen «on demand» konverzijom u DjVu format. To konkretno znači da je slika na serveru pohranjena u JPEG formatu, a konvertira se u DjVu format po zahtjevu korisnika s Interneta, koji sliku želi učitati u svoj web preglednik. Time se znatno smanjuje količina podataka koju je potrebno

[2] Detaljnije o tehničkim standardima vidjeti na: www.eromm.org/standards.htm .

prenijeti mrežom, čime rad korisniku postaje ugodniji i jeftiniji. DjVu je tehnologija sažimanja slike. Razvijena je 1996. u AT&T Laboratories, kako bi riješila problem distribucije slika visoke rezolucije preko Interneta. DjVu postiže 5 do 10 puta bolji stupanj sažimanja od JPEG i GIF formata za dokumente u boji, i od 3 do 8 puta od TIFF formata za crno-bijele dokumente. Skenirana stranica u razlučivosti od 300 DPI, u punoj kolor-skali može biti sažeta na veličinu od 30 do 100 KB s njezine originalne veličine od 25 MB. Crno-bijele stranice na 300 DPI nakon sažimanja zauzimaju od 5 do 30 KB. Da bi neku sliku u DjVu formatu učitali u web - preglednik, moramo instalirati DjVu plug-in, koji je dostupan za različite sistemske platforme. DjVu plug-in omogućuje pomicanje, rotaciju i zumiranje slike. U Knjižnici su provedena opsežna komparativna istraživanja različitih metoda sažimanja[3]. Na temelju rezultata, Knjižnica se odlučila za DjVu metodu.

Nakon skeniranja i vizualne provjere ispravnosti, skenovi se povezuju s metapodatcima. Odabrani naslovi podvrgavaju se postupku optičkoga raspoznavanja znakova (Optical Character Recognition – OCR), kako bi se postigao pretraživi tekstualni oblik. Takav oblik dostupan je korisnicima samo uz alat za pretraživanje koji se zove Retrieval Ware (prije Excalibur). Retrieval Ware pokazao se vrlo dobrim i kod tekstova koji su ili loše digitalizirani ili imaju OCR greške. Slike i rezultati nastali OCR-om povezuju se jer izvorna slika dokumenta osigurava veći stupanj autentičnosti.

3.3. Pristup i pohrana digitalnih zapisa

Za čuvanje, arhiviranje i pristup digitaliziranom gradivu služi sustav Adic Scalar 1000, koji se sastoji od automatizirane knjižnice na magnetskim vrpcama («tape library»), diskovnoga podsustava, datotečnoga sustava SAM FS i programske aplikacije AIP Safe. Adic Scalar 1000 je skalabilan, modularan sustav za pohranu, s naprednim opcijama za zaštitu pohranjenih podataka. Njime upravlja datotečni sustav (file system) Sun SAM-FS (The Sun Storage and Archive Manager File System), instaliran na četveroprocesorsko računalo Sun Mycrosystem 450, koje kontrolira pristup svim pohranjenim datotekama i svim uređajima vezanim na sustav. Takva metoda pohrane podataka naziva se *tape based nearline dana storage*. Nearline dana storage metoda pohrane jeftinija je od online dana storage metode, ali je istodobno i sporije vrijeme dohvata podataka. Sve digitalizirano gradivo sprema se na magnetske vrpce, a metapodatci i najčešće upotrebljavan skup slika na diskove na odvojenom serveru. Evidenciju isteka roka trajanja pojedine magnetske vrpce i migriranje na novu magnetsku vrpcu, sustav obavlja automatski, bez intervencije djelatnika Knjižnice. Svi skenovi pohranjuju se na dvjema magnetskim vrpcama koje se u sustavu nalaze u online statusu, a treća je magnetska vrpca spremljena offline u strogo kontroliranim mikroklimatskim uvjetima.

Knjižnica ima 237 utora (*slotova*) za magnetske vrpce, 18 diskova u diskovnom podsustavu; od toga je još 12 diskova neupotrijebljeno. Kako Knjižnica planira nabavku još jednog uređaja za digitalizaciju (to će biti ili skener za knjige, ili hibridna mikrografska kamera), povećat se i potreba za prostorom, pa se razmišlja o proširenju sustava za još 350 utora za magnetske vrpce.

---

[3] Za detalje vidjeti radove A. Knolla: *Compression of bi-level images* (http://www.nkp.cz/start/knihcin/digit/vav/bi-level/Compression_Bi-level_Images.html) i *Testing new image formats for document delivery* (http://www.nkp.cz/start/knihcin/digit/vav/djvuhybrid/tests/DocDeliveryImageFmt.html) .

Digitalnim dokumentima moguće je pristupiti preko Interneta ili intraneta. U Knjižnici postoji poseban server, čija je zadaća isključivo“on demand” konverzija JPEG-ova u DjVu format.

Knjižnica, osim robotiziranoga sustava pohrane na magnetskim vrpcama, za manje količine podataka rabi metodu pohrane na CD-R medije. Uglavnom se tako pohranjuju zapisi nastali izravnim digitaliziranjem rukopisa i rijetkih spisa. S obzirom da CD-R mediji podliježu degradaciji, Knjižnica je ustanovila sustav kontrole, koji se temelji na mjerenju redundancije medija. Na temelju te metode moguće je odrediti rok u kojem je nužno migrirati podatke na novi medij. Svaki CD-R medij postoji u dvjema kopijama.

4. DOBM SGML

Kad je sredinom 1990-ih Knjižnica započela sa sustavnom digitalizacijom, postalo je razvidno da sve veća zbirka digitaliziranoga gradiva zahtijeva i nekakvu metodu upravljanja njezinim sadržajem. Tad još nije postojao XML, niti rješenje kako opisati i strukturirati digitalni dokument, osim preporuke da bi SGML[4] bio dobra metoda.

Tijekom digitalizacije nastaju dvije grupe podataka: prvu grupu čine slike koje nastaju digitalizacijom mikrofilmova (ili izravnom digitalizacijom rijetkih spisa i rukopisa), a drugu grupu opisi tih slika. Prvu grupu označujemo imenom «podatci», a drugu vrstu «metapodatci». DOBM (Description of Old Books and Manuscripts) razvijen je za potrebe projekta Memoria Mundi Series Bohemica, a 1999. UNESCO ga prihvaća kao međunarodni standard za digitalnu produkciju unutar projekta Memory of the World. DOBM rabi SGML kao podlogu i nezavisan je od bilo koje softverske ili hardverske platforme. Detaljan prikaz strukture DOBM-a Knjižnica je objavila na CD–ROM-u i na Internetu, vidjeti URL:

http://www.nkp.cz/start/knihcin/digit/WWW/doc/digitiz.htm .

Kako je danas XML prihvaćen kao jezik za opisivanje strukture podataka, odlučeno je da se s DOBM SGML-a prijeđe na XML. Konverzija je planirana tijekom ove godine. S tim u svezi, napisan je novi XML-baziran DTD[5].

Izvori:

http://www.leidykla.vu.lt/inetleid/inf-mok/20/str24.html
http://www.unesco.org/webworld/mdm/czech_digitization/doc/digitiz.htm
http://www.nkp.cz/start/knihcin/digit/WWW/doc/digitiz.htm .
http://www.nkp.cz/start/knihcin/digit/vav/bi-level/Compression_Bi-level_Images.html
http://www.nkp.cz/start/knihcin/digit/vav/djvuhybrid/tests/DocDeliveryImageFmt.html
http://djvu.sourceforge.net/abstract.html
http://www.icao.int/djvu/pr/index.html
http://www.w3schools.com/xml/xml_whatis.asp

Renata Horvat

[4] SGML – Standard Generalized Markup Language (ISO 8879:1985) – međunarodni je standard za definiranje aplikacijski i platformski neovisnih metoda za zapis tekstova u elektroničkome obliku.
[5] DTD – Document Type Definition točno navodi oznake elementa koji se mogu rabiti u XML dokumentu i raspored tih elemenata.