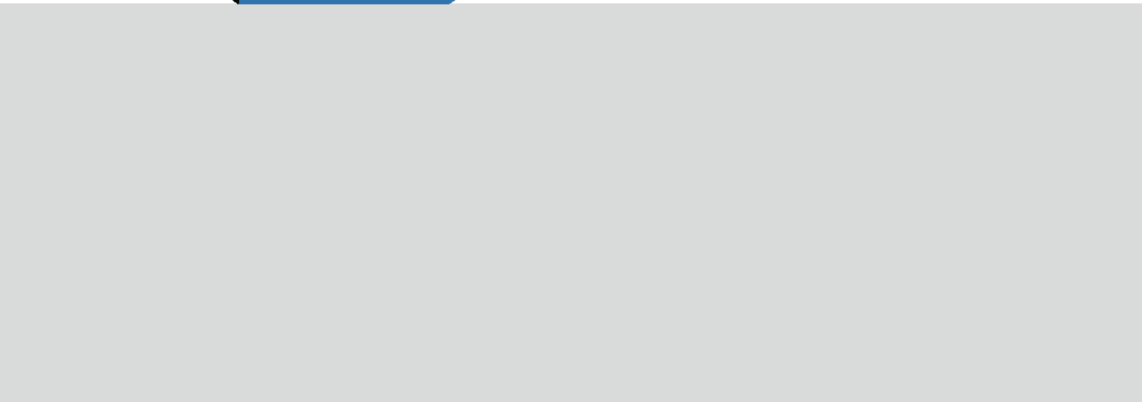


KNIHOVNA
AKADEMIE VĚD ČR
NÁRODNÍ 3, 115 22 PRAHA 1

INFORMACE

1/2008



Jarmila Burgetová

ÚVODEM

Chtěla jsem začít větou „INFORMACE v novém kabátě“, ale to tu už jednou bylo! Při prohlížení starších čísel INFORMACÍ (pro nezasvěcené uvádím, že letos vstupujeme už do 17. ročníku) jsem narazila na článek s tímto názvem, a to v prvním čísle roku 1994. Tehdy jsme začali vydávat INFORMACE s novou, graficky upravenou obálkou, která měla symbolicky naznačit, že chceme nové informace šířit jako „hlásná trouba“ do celého světa. Onen „nový kabát“ nám vydržel dlouhých 14 let, byť celkem 5× změnil svou barvu.

Dnes má knihovna nové logo a s ním korespondující pracovní vizitky, poprvé vydala samostatnou a nově oděnou Výroční zprávu za rok 2007 a teď máme také novou obálku INFORMACÍ. Každého asi hned napadne: proč tolik změn? Ano, máme také nového ředitele...

Ovšem cíl našeho zpravodaje, jak INFORMACE nadále skromně nazýváme, zůstává stále stejný: INFORMACE chtějí být platným pomocníkem knihoven a středisek vědeckých informací Akademie věd ČR – všestranným informátorem o knihovnickém dění jak v Akademii, tak i mimo ni.

Tam, kde jednotlivá čísla našich INFORMACÍ uchovávají, mají skvělou možnost zpětně sledovat nejdůležitější

události, které se staly v průběhu posledních šestnácti let. INFORMACE podávají tedy obraz vývoje KNAV a knihoven AV ČR prakticky od počátku devadesátých let 20. století. Je to dokumentace, kronika i archiv v jednom, mimoděk mapující historii jedné velké specializované knihovny s důležitou sítí odborných knihoven.

Je zajímavé sledovat, jak se obsah INFORMACÍ v průběhu let měnil. Od jednoduchého předávání aktuálních informací (léta 1992–1994), přes pokus o vydávání plnohodnotného časopisu postupně vylepšovaného s řadou stálých rubrik (léta 1995–1999), přes monotematická čísla (z let 2000–2003), až k poslední koncepci, podle níž jsme ve čtyřech pravidelně vydávaných číslech předkládali v prvním čísle výroční zprávu KNAV za předcházející rok, v druhém až třetím čísle (mnohdy spojeném ve dvojčíslo) jsme otiskovali aktuální články o interních aktivitách KNAV a konečně ve čtvrtém čísle byly publikovány články pracovníků KNAV, v nichž zpracovávali své poznatky a zkušenosti získané na zahraničních služebních cestách. V roce 1997 jsme se chtěli rozloučit s papírovou verzí zpravodaje (od druhého čísla roku 1997 je na webu KNAV zároveň souběžně zpřístupňována elektronická verze), avšak na žádost odběratelů jsme pokračovali i v roce 1998 a dalších v dosavadní praxi.

V hledání nových cest budeme pokračovat. V roce 2008 poprvé vyšla výroční zpráva v samostatné publikaci, takže INFORMACE (pokud chceme zachovat periodicitu čtvrtletníku) budou mít více prostoru. Záměr vydávání tematických čísel chceme kombinovat s dosavadní osvědčenou praxí informovat o tom, co nového se děje v KNAV a sdělovat čtenářům a zájemcům to, co nového se odehrálo na zahraničních kongresech, konferencích, seminářích, kterých se pracovníci KNAV zúčastnili. Věřme, že se nám to podaří.

Stěžejním tématem letošního prvního čísla INFORMACÍ je v poslední době soustavně probíraná digitalizace. Zařazené příspěvky by však měly pojednat tuto tematiku na základě našich zkušeností – tedy především informovat o tom, jak vznikalo a vyvíjelo se naše digitalizační centrum v Jenštejně a v druhém případě, jaký přínos bude mít pro bádání v matematických vědách projekt a budování České matematické digitální knihovny.

Těšíme se na vaši odezvu!

ÚVOD

Digitalizace je velmi aktuální téma v celém světě.

Všichni si uvědomují, že záchrana historických a vědeckých dokumentů pro další generace pomocí digitalizace je důležitým mezníkem pro zachování kulturních hodnot. V České republice lze již několik let pozorovat řadu aktivit spojených s digitalizací, které již mají své výsledky v podobě digitálních knihoven s řadou digitalizovaných kulturních a vědeckých materiálů.

Cílem tohoto příspěvku je nastínění vzniku a vývoje Digitalizačního centra Knihovny AV ČR a především jeho pracovních postupů, používaných při digitalizaci vědeckých i historických dokumentů.

VÝCHODISKA PRO VZNIK DIGITALIZAČNÍHO CENTRA

Digitalizační iniciativy na evropské půdě

Evropská digitální knihovna

Jedním z příkladů evropských iniciativ je projekt Evropská digitální knihovna (dále jen EDL). Jedná se o cílený projekt podporovaný Evropskou komisí, který je koordinován Německou národní knihovnou.

Projekt EDL je postaven již na existující Evropské knihovně, jejíž provoz a přístup k elektronickým zdrojům zajišťuje CENL (Conference of European National Librarians).

Vznik projektu EDL je odpovědí na žádost komisařky pro informační vědu a média Viviane Redingové, která zazněla na konferenci CENL v Luxemburgu 29. července 2005.

Realizace projektu EDL byla zahájena v červenci roku 2006. Jeho cílem je integrace a propojení bibliografických katalogů a kolekcí digitálních sbírek národních knihoven evropských států v multijazyčné podobě, ochrana a archivace digitálních zdrojů a podpora digitalizace.

Tento projekt spadá pod skupinu programů eContentplus Programme, které jsou určeny specializovaným okruhům týkajících se především veřejného sektoru, vzdělávání a digitálních knihoven.

Minerva

Do skupiny programů eContentplus Programme patří kupříkladu programy Minerva a MinervaPlus. Minerva je síť zajištěná vládami členských států EU, která se zaměřuje na koordinaci národních digitalizačních programů, na vytváření platných pravidel, doporučení a příruček pro digitalizaci, dlouhodobý přístup a uchovávání dat.

Minerva Plus je od roku 2004 rozšířením a pokračováním programu Minerva. Účastníkem za Českou republiku je Ministerstvo kultury ČR. Od roku 2006 je program Minerva rozšířen o program Minerva eC. Toto konsorcium se věnuje hodnocení digitalizačních aktivit, poskytuje společné investory a experty z celé Evropy, shrnuje výsledky dosažené programem Minerva.

Cílem tohoto programu je vytvářet přidanou hodnotu produktů a služeb na Evropské úrovni, podporovat a rozvíjet podvědomí o národním umění, přispívat k překonání roztržitého a duplikace digitalizačních aktivit v kultuře a vědě a maximalizovat spolupráci členských zemí. Z České republiky je zde zastoupena např. Národní galerie v Praze, Národní knihovna ČR, Národní muzeum, Národní technické muzeum nebo Moravské muzeum.

Iniciativa i2010: Digital Libraries

i2010: Digital Libraries je „vlajkovou lodí“ Evropské komise pro celkový vzestup „digitálních aktivit“. Cílem iniciativy je vytváření různorodého kulturního a vědeckého dědictví, použitelného online pro pracovní potřeby, studium a volný čas. Je postavený na evropském kulturním a vědeckém bohatství. Iniciativa se skládá ze dvou částí: kulturní a vědecké.

Pro realizaci digitálních technologií, k rozšíření a pro snadný přístup k informacím je nutné zajistit tři kroky:

Online přístupnost

předpoklad pro maximální využití vědeckými pracovníky i veřejností

Digitalizace

digitalizovat kolekce publikovaných prací pro jejich širší využití v informační společnosti.

Ochrana a uchování

zabezpečení obsahu prací pro budoucí generace.

V iniciativě i2010: Digital Libraries je Česká republika zastoupena Ministerstvem kultury.

TEL-ME-MOR

Na projekt EDL navazuje projekt TEL-ME-MOR, který je zaměřen na šíření a prezentaci vědeckovýzkumných informací v digitální podobě. Program je určen především novým členským zemím EU.

Evropská knihovna poskytuje prostřednictvím portálu přístupy do jednotlivých národních katalogů a digitálních knihoven. V současnosti jsou přístupné bibliografické katalogy a digitální knihovny z více než 33 států Evropy včetně České republiky. Zástupcem ČR je Národní knihovna ČR s digitální knihovnou Kramerius (bohužel jenom v české verzi) a Manuscriptoriem, souborným katalogem ČR, katalogem Národní knihovny ČR a Českou národní bibliografií.

V Evropské digitální knihovně je zastoupeno například i Srbsko. Již před několika lety vznikla v Srbsku národní iniciativa Národní centrum pro digitalizaci. Na vzniku se společně podílelo několik subjektů:

Matematická fakulta v Bělehradě
Matematický institut Srbské akademie věd
Archeologický institut Srbské akademie věd
Národní knihovna Srbska
Národní muzeum v Bělehradě
Srbský archiv
Ústav na ochranu kulturního dědictví
Jugoslávská filmotéka
Muzeum města Bělehradu

Cílem Národního centra pro digitalizaci je digitalizace studijní matematické literatury, rukopisů, archivních a muzejních materiálů.

V Bulharsku stáli u počátku digitalizace také matematictí vědci – Matematický institut Bulharské akademie věd, stejně tak vědci v Litvě – Matematický institut Litevské akademie věd. V roce 2003 vzniklo v Bulharském městě Borovec digitalizační konsorcium SEEDI – South-Eastern European Digitization Initiative. Do této iniciativy se zapojilo kromě Národního centra pro digitalizaci v Srbsku, v Bulharsku, Rumunsku, Ukrajině dále vědci z Francie, Dánska, Malty, Řecka a Španělska nebo také z Čech. Cílem konsorcia je:

- šířit podvědomí o potřebě digitalizace kulturního a vědeckého dědictví v jihovýchodní Evropě
- lokalizovat a mobilizovat lidské a materiální zdroje v regionech
- zlepšovat komunikaci a šířit informace mezi institucemi s podobnými vědeckými záměry
- napomáhat projektům, které vznikají v jihovýchodní Evropě
- podporovat spolupráci EU a států z jihovýchodní Evropy

Většina dalších států kromě Srbska a Malty, zapojených do iniciativy SEEDI, je v současnosti zastoupena i v Evropské digitální knihovně.

Prostředí digitalizace v ČR

Pro digitalizaci v České republice má zásadní význam program Ministerstva kultury ČR „Veřejné informační služby knihoven“ (dále již jen VISK), který schválila vláda ČR v roce 2000. Hlavním cílem programu je inovace veřejných služeb knihoven na bázi informačních technologií. Koordinátorem programu VISK je Národní knihovna ČR.

Program VISK má celkem devět dílčích částí, které jsou vzájemně provázány.

Pro digitalizaci jsou podstatné především podprogramy VISK 5, VISK 6 a VISK 7.

VISK 5 je národním programem pro retrospektivní konverzi katalogů knihoven v ČR: RETROKON. Jedná se o převod bibliografických záznamů vydávaných dříve v lístkových katalogích do digitální podoby. Tato digitalizace je důležitá především pro přístupnost fondů knihoven místním i vzdálenějším uživatelům. Digitální podoba těchto lístkových katalogů by měla být v ideálním případě textová, strukturovaná dle mezinárodních standardů a v nutném případě může být v podobě obrazu.

Program VISK 5 má dvě etapy plnění, první etapa probíhala v letech 2000–2003 a druhá probíhá v letech 2004–2007.

Druhý program VISK 6 je národním programem pro zpřístupnění vzácných dokumentů pomocí digitalizace a jejich ochrana. Jedná se především o „Memoriae Mundi Series Bohemica“ (dokumenty dochované na území České republiky). Úzce spolupracuje s programem UNESCO Paměť světa, pro který bylo již zpracováno několik pilotních projektů.

Třetí podprogram, VISK 7 neboli Kramerius, je národním programem pro mikrofilmování a digitální zpřístupnění dokumentů ohrožených degradací kyselého papíru. Cílem projektu je záchrana a zpřístupnění národních dokumentů z kyselého papíru. Především u novin je již poškození značné a je nevhodné jejich využívání pro klasickou výpůjční službu.

Realizace projektu spočívá v převedení ohrožených a poničených dokumentů na mikrofilm, který slouží pro archivní účely. V druhé fázi je možné mikrofilm digitalizovat a následně zpracovat obrazy pro potřeby zpřístupnění na lokální síti nebo internetu. Výstupem jsou tři druhy produktů: archivní mikrofilmy, matriční negativy (slouží pro pořízení uživatelských kopií mikrofilmu) a pozitivní mikrofilmy (nahrazují originální dokument, v případě následné digitalizace je uchována digitální kopie). Archivace mikrofilmů vyžaduje speciální podmínky, proto ji zajišťuje Národní knihovna ČR, která disponuje optimálními prostory a podmínkami [Polišenský, 2005]. Jde tedy o nepřímou digitalizaci dokumentu.

V současné době je již možné v rámci programu VISK 7 provádět i digitalizaci přímou, tak jak probíhá např. v Digitalizačním centru Knihovny AV ČR (dále již jen DC KNAV). Je však třeba dodržovat pravidla předepsaná Národní knihovnou ČR a programem VISK 7, tzn. výstupem projektu musí být kromě digitálních dokumentů se standardními parametry i mikrofilm. Instituce mohou čerpat finance po splnění všech podmínek daných programem. Samotnou digitalizaci pro program VISK 7 zajišťují některé komerční firmy, např. Elsynt Engineering s.r.o., Ampaco ČR s.r.o. nebo Microna a mikrografická pracoviště knihoven [Polišenský, 2005].

Většina institucí v ČR však nemá vlastní digitalizační pracoviště a využívá těchto programů pro záchranu svých dokumentů pomocí dodavatelů a spolufinancování digitalizace z vlastních prostředků.

Jednou z knihoven, které využívají k digitalizaci dodavatele, je Jihočeská vědecká knihovna v Českých Budějovicích. Knihovna digitalizuje především periodika poničená při záplavách, starší regionální mapy a periodika, historické dokumenty a mapy. V současnosti má digitalizováno 52 000 stran periodik a 530 kusů map. Periodika pro Jihočeskou vědeckou knihovnu v Českých Budějovicích digitalizovala firma Elsynt Engineering s.r.o. a mapy v loňském roce digitalizovalo Digitalizační centrum KNAV. Digitalizace map zahrnovala i vytvoření metadat k tomuto speciálnímu druhu dokumentu.

Zcela nové vlastní digitalizační pracoviště má v současnosti Městská knihovna v Praze, která plánuje postupnou digitalizaci celého svého fondu, v první řadě však historický a vzácný fond knih vydaných do roku 1860. Z předchozích projektů má knihovna v digitální podobě již 30 titulů knih zpracovaných Národní knihovnou.

Toto nové digitalizační středisko je v provozu od 1. listopadu 2007, je vybaveno dvěma skenery, barevným typu Konica Minolta PS 5000 C pro formáty A3 včetně a černobílým skenerem Zeutschel Omnicam 5000 TT Table Top 90° pro formáty do A2 včetně. Pro grafické úpravy je používán software Sirius.

PŘEDPOKLADY, VZNIK A VÝVOJ DIGITALIZAČNÍHO CENTRA KNIHOVNY AV ČR

Jednou ze základních potřeb pro kvalitní výzkum a vývoj je efektivní přístup k informacím obsaženým v odborné literatuře. Pracoviště Akademie věd ČR vydávají již několik desetiletí velký objem vědecké literatury, jejíž velká část je publikována ve vědeckých časopisech, mnohé tituly mají více než stoletou tradici [Lhoták, 2004].

Ve svých počátcích byly vydávány časopisy ve vědeckých spolcích a společnostech. Ty se pak staly zdrojem fondů Základní knihovny ČSAV vzniklé v roce 1952. Její nástupkyní je

současná Knihovna AV ČR, v. v. i. Podrobná historie Knihovny AV ČR, v. v. i. (dále jen KNAV) je popsána např. v diplomové práci Michaely Novotné **Transformace knihovních služeb na přelomu století: příklad na Knihovně Akademie věd ČR.**

Vznik digitální knihovny na půdě AV ČR byl logickým vyústěním celosvětového trendu a potřeb uživatelů vědecké knihovny.

V současné době vzniká naprostá většina vědeckých prací v elektronické formě. Pro tento typ dokumentů se používá anglický název born-digital. Pro výzkum v řadě oborů jsou však důležité i velmi staré vědecké práce a jejich převedení do digitální podoby pak musí zajišťovat digitalizační pracoviště.

Významným impulsem pro vznik digitalizačního pracoviště v KNAV byly povodně v roce 2002. Pro záchranu obsahu poškozených knih a časopisů byla digitalizace jedním z možných řešení. Při tak velkém objemu zničených publikací nebylo možné všechny restaurovat. Na vznik DC KNAV přispěla i Evropská unie z finančních prostředků určených na pomoc při odstraňování škod po povodních.

Zkušenosti a inspirace pro vybavení digitalizačního pracoviště byly čerpány v evropských knihovnách se zavedenými digitalizačními pracovišti (např. Univerzitní knihovna v německém Göttingenu.). V českém prostředí byla navázána úzká spolupráce s Národní knihovnou ČR, která měla v době vzniku DC KNAV dvě digitalizační pracoviště, jedno pro rukopisy a staré tisky a druhé pro záchranu tisků z 19.–20. století tištěných na kyselém papíře. Vznikla rámcová představa, jak by mělo digitalizační pracoviště KNAV vypadat po technické, personální a organizační stránce. Na základě těchto nabytých poznatků byly vybírání dodavatelé skenovacích zařízení a softwaru pro další zpracování (kapitola č. 2 Vybavení pro DC KNAV). Spolupráce s Národní knihovnou se v následujících letech rozšířila na několik dalších projektů.

V průběhu roku 2003 probíhal výběr hardware a software, konečný výběr byl uskutečněn na konci tohoto roku. Na počátku roku 2004 byli přijati 4 pracovníci, kteří měli za úkol podílet se na vývoji postupů digitalizace a zajistit kvalitní a efektivní produkci. První polovina tohoto roku bylo proto obdobím testování. Byly vyzkoušeny různé metody digitalizace a následného zpracování, postupně se vyprofilovaly nejvhodnější způsoby. Vzhledem k vývoji nových technologií, softwaru a získávání nových poznatků z prostředí digitalizace se průběžně zlepšují postupy směrem k vyšší kvalitě a především k větší efektivitě práce. Digitalizační pracoviště úzce spolupracuje s tvůrci produkčního software Sirius a díky této spolupráci se vlastnosti programu neustále vylepšují. Účast na výzkumných projektech týkajících se problematiky digitalizace má pro provoz digitalizačního centra významný přínos po odborné stránce i z hlediska optimalizace workflow a zvyšování efektivity produkce.

Od počátku digitalizace v roce 2004 do konce listopadu roku 2007 bylo digitalizováno 1 872 504 stran, byl zvýšen počet počítačů a většina z nich byla již několikrát

inovována. Byl pořízen velkokapacitní skener pro jednotlivé listy, upgradován a kapacitně rozšířen produkční server a posílena počítačová síť.

Projekty

Digitalizační centrum KNAV se podílí v současnosti na třech výzkumných projektech – „Digitální knihovna Akademie věd ČR“, „Česká digitální matematická knihovna“ a „Optimalizace nástrojů pro digitalizaci tištěných dokumentů ohrožených degradací kyselého papíru.“

Spolupráce v digitalizaci

Od svého vzniku navázalo DC KNAV spolupráce různorodého charakteru, od černobílého skenování až po specializované úpravy v grafickém softwaru. Za nejzajímavější materiály, které se dostaly ke zpracování v poslední době, je možné označit diplomy a ocenění Václava Havla z jeho knihovny, Šteyerův Kancionál pro Oddělení hudební historie Etnologického ústavu AV ČR nebo barevné mapy pro vědeckou knihovnu v Českých Budějovicích. Pro Umělecko-průmyslové muzeum byly zpracovány od skenování až po metadata a OCR tři tituly odborných časopisů.

V současnosti jsou digitalizovány barevné rukopisy ze 14.–16. století pro Moravskou zemskou knihovnu v Brně. Hotové materiály budou umístěny v digitální knihovně Manuscriptorium (dostupná na <http://www.manuscriptorium.com>). Systém Manuscriptorium je určen pro shromažďování a zpřístupnění informací o rukopisech a starých tiscích. Systém provozuje firma AIP Beroun s.r.o. a je financován Národní knihovnou ČR.

Mezi dlouhodobě spolupracující instituce patří např. Fakulta Sociálních věd Univerzity Karlovy, pro kterou byly skenovány diplomové práce, nebo Filozofická fakulta, Ústav národního korpusu, pro který jsou pravidelně digitalizovány knihy od švédských autorů určené pro jazykový rozbor.

Tento rok byla navázána spolupráce s Ústavem pro jazyk český na digitalizaci Jungmannova slovníku a dalších starých tisků, dále pokračuje digitalizace pro Moravskou zemskou knihovnu v Brně, digitalizace diplomových prací pro Univerzitu Karlovu nebo digitalizace divadelních scénářů pro Divadelní ústav.

TECHNICKÉ VYBAVENÍ A ORGANIZACE PRÁCE V DC KNAV

Hardwarové a softwarové vybavení

Kvalitní technické zabezpečení digitalizačního centra je důležitým faktorem pro efektivní fungování celého pracoviště. Na podzim roku 2003 byly osloveny firmy s poptávkou na dodávku skenerů a softwaru pro následné zpracování. Výběrové řízení na základě ustanovení § 49a zákona č. 199/1994 Sb., o zadávání veřejných zakázek vypsal tehdejší vedení KNAV. Předmětem výběru bylo zakoupení dvou černobílých

skenerů na formáty do rozměru A2 a jednoho barevného skeneru na dokumenty do formátu A1. Požadavky na konkrétní parametry skenerů byly následující:

Požadavky na knižní skener barevný:

- max. velikost dokumentu A1 (rozevřená kniha formátu A2)
- min. rozlišovací schopnost pro formát A1 dosažená fyzicky (nikoliv interpolovaná) 200 dpi
- barevné zobrazení, šedá škála, černobílé zobrazení
- stůl pro vyrovnání rozdílů vazby (1 80°)
- možnost skenování vázaných dokumentů rozevřených pouze na 90 až 120°
- dostatečná vzdálenost mezi dolní polohou stolu a přítlačným sklem, příp. odklopné sklo
- šetrné osvětlení (vůči dokumentu i obsluze)
- možnost ovládání expozice rukou i nožním pedálem

Programové vybavení: možnost ukládat dokumenty ve formátech TIFF, JPEG, GIF, PNG, úprava nasnímaného obrazu (ořezy, narovnání, vyrovnání zaoblení, úprava jasu a kontrastu), dávkové zpracování, editace souborů (změny informací v záhlaví, přejmenování), základní indexace obrázků.

Samozřejmým požadavkem bylo uvedení ceny servisu včetně náhradních dílů, update a upgrade softwarového vybavení. Důležité bylo také uvedení životnosti CCD prvků, úroveň hloubky barev, požadavky na kalibraci zařízení a jeho předepsané servisní úkony [Lhoták, 2003].

Požadavky na knižní skener černobílý:

- max. velikost dokumentu A2 (rozevřená kniha formátu A3)
- min. rozlišovací schopnost pro formát A2 dosažená fyzicky (nikoliv interpolovaná) 200 dpi
- šedá škála, černobílé zobrazení
- stůl pro vyrovnání rozdílů vazby (180°)
- možnost skenování vázaných dokumentů rozevřených pouze na 90 až 120°
- případná možnost použití přítlačného skla
- šetrné osvětlení (vůči dokumentu i obsluze)
- možnost ručního ovládání expozice i nožním pedálem

Požadavky na programové vybavení, jeho funkce a servisu byly stejné jako u skeneru barevného.

V případě barevného velkoformátového skeneru byly osloveny české firmy Ampaco ČR s.r.o., MIKROGRAFIE s.r.o, AIP Safe s.r.o., slovenská BAUKIS s.r.o. a Flow Plus s.r.o. a německá CRUSE GmbH. V případě černobílých skenerů byla kromě výše uvedených společností oslovena ještě firma MINOLTA s.r.o. Firmy byly vyzvány k podání nabídek

na dodávky skenovacích zařízení dle výše uvedených parametrů a dalších dodatečných podmínek obsažených v návrhu veřejné zakázky.

Výběrová komise doporučila barevné zařízení DigiBook a černobílé skenery Zeutschel. Skener DigiBook RGB 10000 od francouzské firmy I2S dodala firma AIP Safe s.r.o. (<http://www.aipsafe.cz/>), jeho servis zajišťuje firma XEROX Czech republic s.r.o. (<http://www.xerox.cz/cs/>). Černobílé skenery od německé firmy Zeutschel GmbH (<http://www.zeutschel.de/>) dodala firma Ampaco ČR s.r.o. (<http://www.ampacocr.cz/>), která zajišťuje i servis. Oba modely splnily podmínky výběru.

Barevný knižní skener DigiBook 10 000 RGB je skenerem pro formáty do A1 včetně při rozlišení 400 dpi. Maximální rozlišení skeneru je 800 dpi při 1 × A4. Obsahuje CCD prvky 3 × 10 000 pixelů RGB. Váha knihy může být až 50 kg a její tloušťka 50 cm. Skener snímá předlohy buď barevně, černobíle nebo ve 256 odstínech šedi, pracuje rychlostí cca 120 obrázků za hodinu při formátu A5 a rozlišení 400 dpi.

Ke skeneru bylo dodáno speciální úchytné zařízení s přítlačným sklem pro knihy, které nelze otevřít více než na 90° nebo 120°. Dalším příslušenstvím byla přídatná plochá deska na skenování velkých map a plochých neknižních předloh. Později bylo pořízeno ještě přítlačné sklo na klasickou vyrovnávací kolébku pro formáty do A3 včetně. Kolébka je pevnou součástí skeneru, kterou nelze demontovat. Součástí dodávky skeneru byl i počítač a software pro ovládání skeneru.

Nastavení skeneru vyžaduje odborné proškolení, kterým musí projít obsluha skeneru. Vhodné jsou zkušenosti z klasického fotografování. Je nutné ruční ostření a nastavení clony, software si na základě těchto nastavení spočítá dobu expozice. Kalibrace barev se provádí dle bílé a černé předlohy. Pro skenování je vhodné mít skener umístěn v prostředí, kde není přímé a ostré světlo či sluneční záření. Např. světla umístěná nad skenerem mohou značně ovlivnit výsledek skenování.

Samotné skenování není obtížné, provádí se v sedě, k dispozici jsou dva nášlapné pedály, jeden pro povel ke skenování a druhý pro vyrovnávání kolébky, na které kniha leží. Při skenování projíždí celé rameno s kamerou i osvětlením nad předlohou, výbojkové trubice ji rovnoměrně osvětlují a kamera snímá. Následně je obraz automaticky uložen do určeného adresáře. Doba ukládání obrazu přímo úměrně roste spolu s nárůstem fyzické velikosti předlohy a rozlišení. Osvětlení trubic lze regulovat, k dispozici jsou čtyři varianty osvětlení. V praxi se osvědčilo nejsilnější osvětlení. Výjimečně se světlo zeslabuje, v případech kdy se skenuje lesklá předloha.

Černobílé skenery Zeutschel oS 7 000 (Obr. 2 Zeutschel OS 7 000) jsou určeny pro formáty předloh do A2 při rozlišení 400 dpi a 800 dpi při formátu 1× A4. Skener obsahuje CCD prvek o 7 500 pixelech. Tloušťka knihy může být až 50 cm. Skener pracuje rychlostí až 250 obrázků za hodinu při formátu A5 a rozlišení 300 dpi. Součástí skenerů bylo již od samého počátku přítlačné sklo pro celou plochu skeneru.

Sklo lze uchytit v případě potřeby na vertikální opěrné rameno a je tak možné skenovat bez něj. Součástí dodávky firmy Ampaco ČR nebyly počítače, pouze software pro ovládání skeneru. Počítače byly dokoupeny na základě požadované specifikace firmy Zeutschel GmbH od Dell Computer s.r.o. (<http://www.dell.cz>), která dodává počítačové stanice do celé KNAV.



Obr. 1 Skener DigiBook 10 000 RGB



Obr. 2 Skener Zeutschel OS 7 000

Skener má automatické ostření i expozici, uživatel pouze vybírá z možností rozlišení, zesvětlení či ztmavení snímání. V případě nutnosti lze také ručně doostřit a otevřít či zavřít clonu. Kamera je pevně připevněna na vertikálním rameni a při procesu snímání se nehýbe. Pohybuje se pouze určitá část ukrytá ve skenovací hlavě.

Skenování lze spouštět stisknutím klávesy na počítači nebo přídavných nožním pedálem. Přítlačné sklo je možné ovládat ručně, pomocí pedálů nebo poloautomaticky, kdy se sklo automaticky samo zvedne po naskenování obrázku a obsluha ho pomocí pedálu přitlačí zpět k předloze. Předloha je osvětlována postranními zářivkovými trubnicemi. Světlo je stálé a nelze jej regulovat, jako je tomu u skeneru DigiBook.

Čtvrtým skenovacím strojem je Panasonic KV – S1025C High Speed Scanner – pořízený v roce 2007. Jedná se o velkokapacitní skener pro předlohy do A4 včetně. Skenuje v barvě (24bit), šedé škále nebo černobíle. Je určen pro skenování jednotlivých listů, nejedná se tedy o knižní skener. Byl pořízen proto, že některé časopisy či knihy lze rozvázat nebo existují předlohy, které svázané nejsou vůbec. Panasonic dokáže skenovat v rozlišení maximálně 600 dpi při formátu A4, jeho výhoda spočívá ve vysoké rychlosti ve srovnání s ostatním zařízením na pracovišti. Je schopen naskenovat cca 1 800 obrázků za hodinu. Náklady na pořízení jednoho digitálního obrazu strany se s tímto zařízením snižují až o osmdesát procent ve srovnání s planetárními knižními skenery.

Ostatní vybavení, pracovní stanice a server, dodala firma Dell Computer s.r.o. Na zálohování digitalizovaného materiálu byla pořízena také pásková mechanika HP Ultrium II.

Po stránce softwarového vybavení se autoři digitalizačního pracoviště nechali inspirovat Národní knihovnou. Byl pořízen produkční systém Sirius produkt firmy Elsys Engineering s.r.o., který slouží ke grafickým úpravám a popisu dokumentů. Spolu se skenerem DigiBook byl od téže firmy zakoupen program Book Restorer™ pro grafické úpravy. Dále pracoviště disponuje volně dostupným softwarem Irfan View a programem pro úpravu metadat XMetal.

PROCESY DIGITALIZACE

Proces digitalizace dokumentů zahrnuje několik kroků, kterými musí dokumenty projít, než je lze poskytnout uživateli. Tyto procesy vyvinuli sami pracovníci v digitalizačním centru KNAV a jsou specifické pro jejich práci. Během testovacího období digitalizačního centra navštívili pracovníci digitalizačního centra některé instituce, které měly již s digitalizací zkušenosti, např. oddělení digitalizace Národní knihovny nebo digitalizačního pracoviště v německém Göttingenu. Cílem bylo načerpat poznatky, které by pomohly při vývoji dalších postupů. V následujících kapitolách budou popsány jednotlivé kroky, kterými musí dokument projít, než je zveřejněn v digitalizované podobě.

Výběr dokumentů pro digitalizaci

Pro projekt Digitální knihovna AV ČR. byla v první fázi určena především všechna periodika, která kdy vydávala a některé stále vydává Akademie věd ČR a všechny její předchůdkyně. Tato periodika jsou vybírána dle předem vybraného seznamu. Vždy je digitalizován celý titul daného periodika, který je dostupný v KNAV, až do roku 1994 včetně. Po tomto roce je zde předpoklad, že dosud vydávaná periodika mají svoji elektronickou verzi již při svém vzniku. Zatím tyto elektronické verze nejsou součástí digitální knihovny, ale byla již zahájena spolupráce s redakcemi akademických ústavů na převzetí těchto elektronických verzí. V případě, že chybí některý výtisk určený k digitalizaci, je podána žádost na meziknihovní výpůjční službu o zapůjčení tisku z jiné knihovny či ústavu.

V druhé fázi digitalizace budou zpracovány monografie vydané Akademií věd, jejími předchůdkyněmi a jednotlivými akademickými ústavy. Tato druhá fáze digitalizace bude započata ještě před koncem roku 2007.

Kromě těchto naplánovaných fází jsou digitalizované např. staré tisky pro Knihovědné oddělení KNAV a jiné akademické ústavy.

Skenování

Snímání předloh probíhá na všech třech skenerech, které má DC KNAV k dispozici. Na černobílých skenerech Zeutschel probíhá veškeré skenování jak černobílé, tak i v odstínech šedi (osmibitové a čtyřbitové skenování) do formátu A2. Na skeneru DigiBook jsou skenovány všechny barevné předlohy a předlohy větší než formát A2.

První tři roky provozu DC KNAV bylo skenováno na všech třech skenerech po celou standardní pracovní dobu, pět dní v týdnu. Vzhledem k nárůstu objemu nasnímaných dat akademického materiálu a objemu speciálních požadavků na skenování barevných předloh bylo pozastaveno skenování menších formátů do A2 v šedé škále na barevném skeneru DigiBook. (tj. periodik a monografií, naplánovaných v rámci projektu DK AV ČR). Tuto práci plně zastávají dva skenery Zeutschel.

Produkce na obou skenerech Zeutschel dosahuje denně při osmihodinové pracovní době v optimálním případě až 4 000 naskenovaných stran při formátu A5, rozlišení 300 dpi a barevné hloubce čtyř bitů. Je to plně dostačující pro navazující zpracování, které, jak bylo zmíněno, je časově náročnější než skenování. Proto byl také od počátku roku 2007 vyčleněn barevný skener pouze pro barevné, vzácné a velkoformátové předlohy.

Parametry skenování

Běžné dokumenty jsou pro potřeby Digitální knihovny AV ČR standardně skenované v barevné hloubce čtyř bitů (čtyři bity jsou z důvodu toho, že jednobitové skenování nevypadá ve struktuře písma zcela dobře, písmo bývá neostré a při nedokonalosti tisku málo zřetelné) a rozlišení 300 dpi. Všechny obrazy se skenují bez černých okrajů, prakticky jen výřez textu (tak, aby zůstal celý text na straně i s číslem strany, nesmí samozřejmě nic chybět) s bílým krajem tak velkým, aby bylo ještě následně možné podle potřeby upravit jeho šířku.

Staré tisky a jiné vzácné dokumenty jsou skenovány barevně, v rozlišení 600 dpi a většinou s malým, cca tří milimetrovým černým okrajem. V následující tabulce (Tabulka 1 Formáty skenování, barevná hloubka a rozlišení) jsou rozděleny formáty, možnosti jejich barevné hloubky a rozlišení, jak jsou standardně nabízeny ústavům AV ČR a kterými se řídí pracovníci digitalizačního centra.

Formát	černobílá	16 odstínů šedi	256 odstínů šedi	barva
A1	300 dpi	–	300 dpi	300 dpi
A2	300 dpi	300 dpi	300 dpi	300 dpi
A3	300/600 dpi	300/600 dpi	300/600 dpi	300/600 dpi
A4–A5	300/600 dpi	300/600 dpi	300/600 dpi	300/600 dpi

Tabulka 1: Formáty skenování, barevná hloubka a rozlišení

Skenování předloh na rozlišení 600 dpi je kvalitnější, ale dvojnásobně časově náročnější a nákladnější než skenování na rozlišení 300 dpi. Proto se provádí jen u speciálních tisků, starých tisků nebo podle individuálních požadavků.

Při skenování všech druhů materiálů se ukládají obrázky v nekomprimovaném formátu TIFF. Následně je z nich provedena bezztrátová komprese LZW (Lempel-Ziv-Welch). Na počátku provozu DC KNAV byly testovány i jiné formáty, které by mohly být vhodné. Např. formát JPEG byl nadějným kandidátem, ale vzhledem k jeho ztrátové kompresi se neosvědčil. Na konci celého procesu zpracování ztratily obrázky viditelně na kvalitě. Pro čtyřbitové obrázky textů tedy tento testovaný formát není příliš vhodný. Někdy je ale využíván pro barevné obrázky, kde se neobjevuje mnoho textu. Výhodou je, že jeho kompresní poměr u barevných předloh je větší než komprese LZW a ušetří se tak místo na nosičích. Velikost souborů naskenovaných předloh je různá dle rozlišení, barevné hloubky a fyzické velikosti předlohy. V následujících dvou tabulkách (Tabulka 2 Velikost souborů při rozlišení 300 dpi a Tabulka 3 Velikost souborů při rozlišení 600 dpi) jsou rozepsané průměrné velikosti souborů před kompresí LZW při různých formátech, rozlišení a barevné hloubce.

Formát	černobílá	16 odstínů šedi	256 odstínů šedi	barva 24 bit
A1	550 KB	–	67 MB	230 MB
A2	250 KB	18 MB	33 MB	100 MB
A3	120 KB	9 MB	17 MB	50 MB
A4	50 KB	4 MB	8 MB	25 MB
A5	25 KB	2,5 MB	5 MB	15 MB

Tabulka 2 Velikost souborů při rozlišení 300 dpi

Formát	černobílá	16 odstínů šedi	256 odstínů šedi	barva 24 bit
A3	4 MB	35 MB	67 MB	200 MB
A4	2 MB	15 MB	33 MB	100 MB
A5	1 MB	8 MB	16 MB	50 MB

Tabulka 3 Velikost souborů při rozlišení 600 dpi

LZW komprese tyto hodnoty sníží cca o 80% u obrázků v odstínech šedi nebo u černobílých. U barevných obrázků se jejich velikost prakticky nezmění, proto je lepší jejich převedení do formátu JPEG. Komprese a konverze do jiných formátů se provádí v programu IrfanView.

Jak zřejmé z tabulek, se zvyšujícím se formátem či rozlišením se zvětšuje objem dat, proto se také prodlužuje čas skenování. Jestliže u formátu A5, v rozlišení 300 dpi a čtyřbitové barevné hloubce lze nasnímat průměrně až 180 obrázků za hodinu, tak

v případě formátu A1 v barvě je to již jen maximálně cca 25 kusů. Kromě toho se prodlužují veškeré další práce spojené s kontrolou a přenosem dat.

Kromě těchto parametrů skenování získá každá stránka skenovaného periodika svůj originální název. Ten je jedinečný a později z něj lze získat informaci, kde se přesně dokument nachází. Tyto názvy souborů jsou typické pro DC KNAV. Mají následující syntaxi:

Sigla – prvních šest míst názvu je rezervováno pro siglu instituce, ve které je fyzický dokument uložen a byl z ní použit k digitalizaci

ISSN – osm následujících míst po sigle pro ISSN bez pomlčky daného titulu

Rok – čtyři místa pro rok vydání ročníku daného titulu

Příloha – jedno místo pro skutečnost, že výtisk bude přílohou. V případě, že není výtisk přílohou, je zde číslice 0. V případě, že výtisk je přílohou, je zde číslice 1.

Výtisk – čtyři místa pro číslo výtisku daného ročníku titulu

Pořadové číslo souboru – čtyři místa pro číslo souboru, skenovací zařízení je vzestupně generuje během skenování

Hotový název souboru pak vypadá například takto:

ABA007023166251985000310008.tif

Obsah názvu je následující:

Osmý soubor z přílohy třetího výtisku z ročníku titulu Právníka, vydaného roku 1985.

Názvy jsou pro potřebu DC KNAV vyhovující a pokrývají téměř všechny rozmanitosti. Sporné situace se řeší individuálně.

Grafický editor IrfanView

IrfanView je malý a velmi obratný grafický prohlížeč a editor pro prostředí Windows. Jeho tvůrcem je student z Bosny Irfan Skiljan. Program je nekomerční a volně dostupný (<http://www.irfanview.com>). Je vhodným nástrojem jak pro začátečníky, tak i pro profesionály. Má řadu svých vlastních funkcí, které jsou pro něj charakteristické. Byl to např. první program s celosvětovým rozsahem, který podporoval animované obrázky ve formátu GIF a multistránkové ve formátu TIFF.

Software podporuje mnoho grafických formátů nejen obrázků, ale dokáže i přehrát multimediální formáty. Umí například provádět ořezy obrázků, úpravu jejich kontrastu a světlosti či zmenšování jejich velikosti dle různých parametrů.

Pro práci DC KNAV je nejdůležitější funkcí konverze do jiných formátů, komprese a změna hloubky barev. Komprese LZW se provádí v tomto programu vždy po naskenování celého svazku. Dále se již zpracovávají a archivují jen zkomprimované soubory. Změna hloubky barev na 16 odstínů šedi se provádí po úpravách z Bookrestoreru, protože tento software umí obrázky exportovat jen ve 256 odstínech šedi a soubory jsou pro další zpracování příliš velké.

Grafický software Book Restorer™

Nejpodstatnější úpravy obrázků probíhají ve specializovaném softwaru Book Restorer™. Je produktem francouzské firmy I2S, která se specializuje na vývoj, výrobu a prodej digitalizačních systémů pro staré nebo poškozené tisky, software na grafické úpravy digitalizovaného materiálu, kontrolní, diagnostické a třídící systémy například tiskových vad apod.

Do DC KNAV byl pořízen již při vzniku centra spolu se skenerem DigiBook, který je též produktem firmy I2S. Z počátku k tomuto programu neměli pracovníci DC plnou důvěru. Jedním důvodem byl konkurenční software Sirius, který měl teoreticky zvládnout většinu operací, které měl Book Restorer™ podle dokumentace také provádět. Kromě toho je Sirius koncipován tak, že v něm lze vytvářet metadata až na úroveň jednotlivých stránek. Druhým důvodem počáteční zdrženlivosti vůči Book Restoreru™ byl fakt, že komunikoval jen v anglickém jazyce a průvodní dokumentace byla také pouze v angličtině, ani uživatelské rozhraní se na první pohled nezdálo být příliš přívětivé. Nicméně po zjištění, že zdánlivě nadějný Sirius není plně funkční a jeho grafické funkce nepracují jak by měly, byl Book Restorer™ prostudován, vyzkoušen a začleněn do linky zpracování digitalizovaného materiálu. Stal se velmi účinným nástrojem při každodenních grafických úpravách nasnímaných publikací.

V prvním roce působení digitalizačního pracoviště byla k dispozici pouze jedna licence na tento software. Po zjištění jeho kvalit a příchodu nových pracovních sil byla pořízena další licence

Základní funkce Book Restoreru™

Každá programová aplikace vyžaduje určitou konfiguraci stanice, na které by měla bez problémů fungovat. Minimální požadovaná konfigurace pro Book Restorer™ je následující:

jeden nebo dva procesory PENTIUM III
RAM 256 MB
grafická karta – 16 mil. barev
rozlišení monitoru 1280 × 1024 pixelů
systém WINDOWS NT 4.00 service Pack 4 nebo vyšší

Kromě toho je doporučované minimální volné místo na disku cca 4 GB. Ze zkušenosti lze však doporučit velikost paměti minimálně 512 MB, ještě lépe však 1 GB a velikost

disku v řádu desítek GB. Aplikace si ve skutečnosti při zpracování vytváří velké množství pomocných souborů a ty potřebují dost prostoru na disku. Po dokončení zpracování je lze odstranit.

Book Restorer™ je programem pro úpravy, opravy, změny a zpracování digitalizovaných knih. Poskytuje výkonné nástroje pro dosažení vysoce kvalitních výsledků bez ohledu na to, zda s ním pracuje začátečník nebo expert. Významnou vlastností tohoto softwaru je fakt, že dokáže pracovat s velkým množstvím obrázků najednou, resp. umí pracovat dávkově.

Princip zpracování není příliš složitý – celá kniha nebo časopis (resp. celý adresář s naskenovanými stránkami) se naimportuje do programu a operace se pak provádějí buď dávkově automaticky na všech stránkách nebo ručně po jednotlivých stránkách či jednotlivých blocích. Architektura zpracovávané úlohy v Book Restoreru™ je následující – zpracovávaná úloha neboli kniha je rozdělena do sekcí. Pod každou z nich je jedna strana výtisku. Sekce však může obsahovat i více stran a může být chápána např. jako kapitola. Kromě toho může každá stránka obsahovat více obrázků, tzv. bloků. Pro každou stránku i pro každý blok je k dispozici velká nabídka grafických nástrojů pro jejich úpravu.

Jednotlivé grafické nástroje lze organizovat do tzv. skriptů, které je možné uložit pro případ, že by bylo nutné dělat stejné úpravy na jiných obrázcích. V případě DC KNAV se tyto skripty hromadně nepoužívají, jen ojediněle a většinou jen pro jeden výtisk. Přestože se některé funkce, například ořez a geometrická korekce, používají vždy na všech dokumentech a skripty by zde mohly najít své uplatnění, používá každý operátor svou vlastní organizaci práce a jeho rozhodnutí, zda si skripty vytvoří. Také pokud se jedná o naskenované tisky, nejsou si všechny podobné. I když se používají některé stejné operace, nastavují se velmi často pro každou knihu různé parametry.

Další velmi praktickou pomůckou je doprovodný program, který je instalován jako součást balíčku spolu s Book Restorerem™. Jmenuje se The Book Restorer™ Automaton. Jde o aplikaci, která automaticky a na pozadí zpracuje zadané úlohy, aniž musela být spuštěna samotná aplikace Book Restorer™. Ušetří to značné množství času. Operátor si připraví v Book Restoreru™ jednotlivé tisky a navolí si k nim příslušné operace, které je třeba provést. Pak si v Automatonu jen nastaví čas, kdy se má kniha zpracovat. V DC KNAV bývá nastaveno zpracování především na noční hodiny.

Vstupní data pro Book Restorer™ musí splňovat některá formátová a barevná omezení. Program pracuje s formáty JPEG, TIFF (monopage nebo multipage), PDF (monopage nebo multipage) a PNG. Nepodporuje však model CMYK a 16 bitový mód RGB. V digitalizačním centru jsou používány ke zpracování obrázky v barevné hloubce 4 bitů, formátu TIFF s LZW kompresí. Aplikace sice umí pracovat se 4 bitovými obrázky, ale neumí je opět v této barevné hloubce uložit, proto jsou následné výstupy ve 256 odstínech šedi.

Grafické moduly

Uživatelům je k dispozici široká škála možností grafických úprav obrázků. V DC KNAV jsou dávkově využívány funkce geometrická korekce a ořez. Ostatní jsou využívány individuálně na jednotlivé stránky dle nutnosti a potřeby.

Modul Geometrická korekce

Geometrická korekce je funkcí pro vyrovnání a opravu nerovností na stránce.

Nerovnosti vznikají při skenování nebo z důvodu nekvalitní fyzické předlohy.

Při úpravě používá několik různých linií stránky dle uživatelského nastavení:

Detekce okrajů stránky – nalezení černých okrajů stránky (pokud jsou tam zanechány během skenování, v DC KNAV se většinou skenuje bez černých okrajů, jenom výřez textu).

Lineární detekce – zjištění všech podstatných horizontálních a vertikálních čárových složek obrázku. Jde například o rámování obrázků na stránce nebo oddělovače odstavců či sloupců.

Detekce řádků – zachytí řádky textu, pokud jsou dostatečně dlouhé.

Detekce zkosení – zjistí, zda je stránka nějakým způsobem zkosená, tato funkce funguje podobně jako modul **Deskew**.

Kromě těchto kritérií lze nastavit ještě dodatečnou konfigurací modulu položku **Relevance**, která udává hodnotu, jak moc má být výsledek provedení jednotlivých nastavení úspěšný. Při různé kombinaci všech nastavení lze dosáhnout velmi dobrých výsledků. Nicméně je nezbytné, pokud je používán tento nástroj dávkově, obrázky po zpracování zkontrolovat. Občas se stává, že původně relativně rovnou stránku geometrická korekce nejenže nevylepší, ale naopak výrazně zkřiví. Například mohou být v nějaké části stránky vertikálně posunuty řádky a stránka pak vypadá, jako by byla zvlhčená, nebo se špatně detekuje odstavec a celý blok textu je po provedení funkce křivý.

Modul Ořez

Modul Ořez umožňuje řezat obrázky třemi různými způsoby:

rozdělení – rozřezání na několik částí

manuální ořez – podle nadefinovaných okrajů

automatický ořez

První typ ořezu tzv. **Division** neboli rozdělení na části se používá v případě, kdy je třeba rozdělit jeden obrázek na několik částí, např. když má operátor nasnímané dvojstrany a chce je rozdělit na jednotlivé stránky. V prvním kroku si navolí, zda chce řezat vertikálně nebo horizontálně (nelze najednou) a na kolik částí. V druhém kroku je nutno nastavit pozice jednotlivých dělících os. Tyto osy jsou nastaveny pevně

v milimetrech. V případě, že mají části, které je třeba od sebe oddělit, na stránce trochu jinou pozici (nejsou všechny řezané předlohy stejně naskenované), může dojít k tomu, že se stránky nerozříznou přesně podle představ (pokud se ořez pustí dávkově na více obrázků). Tento typ ořezu je využíván v digitalizačním centru jen výjimečně, pokud to povaha naskenovaných předloh vyžaduje.

Druhým typem je Manuální ořez. Zde lze oříznout okraje jednotlivých stránek podle uživatelem pevně nastaveného rámce. V prvním kroku si zpracovatel nadefinuje horní, dolní, pravou a levou osu v milimetrech, o které chce obrázek oříznout. Opět se osy nastavují pevně, proto se tato funkce ani ta předchozí příliš nehodí pro dávkové spouštění.

Naopak velice vhodným nástrojem pro dávkové zpracování je třetí typ ořezu: automatický. Funguje na principu vyhledání informací pro ořez přímo v obrázku. V první volbě si uživatel zvolí typ vyhledávání podle typu stránky, které chce ořezávat. Může zvolit jednoduché vyhledávání – to se aplikuje pouze v případě, že je k dispozici pro ořez jen jedna strana. Druhou možností je volba zpracování více stran najednou: je zde mód pro jednostrany a dvojstrany. V druhém kroku se nastavují okraje pro vyhledávání textu a okraje – uživatel může ořezávat od textu ven, od poloviny mezi textem a okrajem nebo od okraje ven. Dále je nutné nastavit hodnoty, o které se má obrázek oříznout na všech čtyřech stranách, hodnotu citlivosti (co má program považovat ještě za text a co už ne) a volitelnou položku Erase Fingers, která odstraní případně naskenované prsty z obrázku.

V DC KNAV je používán automatický ořez pro více jednostran. Vždy se nastavuje ořezávání tak, aby byl odstraněn zbytečný bílý okraj, zhruba jeden centimetr na horním a dolním okraji a pět milimetrů po levé a pravé straně. Není povinné ořezávat konkrétně jeden centimetr, záleží to na povaze předlohy. Některé předlohy mají tisk téměř do kraje, takže tam zbývá např. jen pár milimetrů volného kraje. Ořez je pak nastaven tak, aby zbyl na levé i pravé straně zhruba stejně široký volný bílý okraj. Špatně ořezané stránky jsou během kontroly individuálně opraveny manuálně.

Modul Binarizace

Binarizace změní obrázek, který je v šedé škále nebo barvě, na černobílý. Pomocí parametrů nastavení hloubky a detailů lze odstranit z obrázku některé nečistoty. V kombinaci s dalšími moduly lze dosáhnout velice uspokojivého výsledku. Obrázky jsou však po této úpravě pouze ve dvou bitech.

Modul Despeckle

Tato funkce je praktickým doplňkem binarizace a je s ní také spjata v tom smyslu, že nelze provést modul **despeckle** bez předchozí binarizace. **Despeckle** odstraňuje buď černé nečistoty na stránce nebo vyplňuje bílá místa neboli chyby v písmu. Napravit lze chyby od velikosti 1 × 1 pixel až do velikosti 24 × 24 pixelů. V DC KNAV se odstraňují

nečistoty maximálně do velikosti 3 × 3 pixelů. U vyšších hodnot se stává, že se začne ztrácet diakritika. Úspěšnost je závislá hlavně na velikosti písma a kvalitě tisku.

Modul Filter

Tento speciální modul zjemní přechod barev a kontur a následně ztmaví písmo. Výrazné zlepšení lze pozorovat např. na fotografiích naskenovaných v šestnácti odstínech šedi, kdy jsou velice špatně rozlišitelné. Filter pomůže k tomu, že se hrubá zrna obrazu slijí a obrázek se tak zjemní a vynikne. Pomocí různých nastavení tohoto modulu se naopak mohou kontury i více zvýraznit. Tento modul lze také použít jenom pro text, kde jsou případně nějaké nečistoty nebo nepříliš výrazné písmo. Filter písmo ztmaví, nečistoty zjemní a když se po té provede binarizace, mohou někdy zcela vymizet.

Modul Colorometric curves

Podstata toho modulu spočívá v tom, že se může změnit paleta odstínů barev v obrázku. Když neodpovídá barevně obrázek skutečnosti, lze pomocí křivky posunout barevnou paletu např. blíže k černé (v případě šedé škály barev). V praxi se tento modul používá v případě, že je stránka například moc tmavá a je na ni vidět špatně text. Pomocí křivky se zesvětlí a text je potom více kontrastní. V kombinaci s binarizací nebo filtrem pak může vzniknout více kontrastní a lépe čitelný text.

Další moduly už nejsou v DC KNAV tak často využívány. Budou představeny pouze ve zkratce:

Area detection – vyhledá na stránce jednotlivé objekty, na které lze pak samostatně provést ostatní grafické úpravy

Adjustment of histogram – pomůže upravit světelné rozpory v obrázku oproti jeho fyzické předloze

Deskew – je možnou alternativou geometrické korekci, ale funguje na jiném principu – stránky vyrovnává jako celek, nikoliv jeho části

Finger masking – tento modul je schopen odstranit z obrázku naskenované prsty a jejich místo vyplnit stejnou barvou, jako je jejich okolí

Light & contrast – pomůže upravit v obrázku světlo a kontrast

Lightning correction – odstraní ze stránek tmavé stíny, které mohou vzniknout např. při skenování

Transformation – slouží k převrácení stránek, pokud požadujeme aby se změnila např. poloha některého objektu na stránce nebo celé stránky

Mezi prakticky nevyužívané moduly v digitalizačním centru patří Negativ – převrácení obrázku do negativu, Polarity Detection – převrácení negativu zpět do pozitivu, Resize – změna velikosti obrázku, OCR – vyčítání textu z obrázku – není pořízena licence pro tento modul, OCR probíhá v jiném programu.

Produkční systém Sirius

Sirius je softwarem vyvinutým firmou Elsys Engineering ve Vyškově (<http://www.ee.cz>). Slouží pro grafické zpracování, tvorbu metadat a archivaci digitalizovaných dokumentů. Může být také samotným softwarem pro komunikaci skener – počítač (pro samotné skenování). Systém pracuje na principu klient – server, proto může být instalován jak v malých kancelářích tak i ve velkých provozech o několika desítkách počítačů.

Systém využívá databázi Microsoft SQL. V databázi jsou uloženy také veškeré indexy, které uživatel během zpracování dokumentům přiřadí a systémové informace o dokumentech. Kromě databáze je zde ještě samotné úložiště dokumentů, které jsou během zpracování ukládány. Pro toto úložiště je vhodné mít vyčleněný dostatečný prostor na serveru, na kterém bylo úložiště vytvořeno. Málo místa pro úložiště dokumentů způsobuje nemalé problémy ve funkčnosti celého Siria. Všechny pracovní stanice, kde je úmysl zpracovávat v Siriovi, tak musí být pomocí počítačové sítě připojeny k serveru. Jinak je Sirius na nepřipojené stanici k síti nefunkční.

Systém Sirius je modulárním systémem a záleží na provozovateli tohoto systému, jaké moduly chce využívat. Software je stále vyvíjen a každé pracoviště má jiné požadavky na funkcionalitu. Vzhledem ke složitosti celého systému budou představeny funkce a moduly využívané v DC KNAV.

DC KNAV využívá Sirius především pro tvorbu metadat, OCR a archivaci. Pro grafické operace využívá DC KNAV především systém Bookrestorer™, který byl již zmíněn výše.

Sirius umí pracovat s těmito formáty obrázků: TIFF, JPEG, BMP, GIF, PNG, MDC, PCD a TGM. Do Siria lze vložit ale i dokumenty např. v PDF, DOC, XLS apod. Není možné je v něm editovat, ale při práci s nimi se spustí příslušný asociovaný program. Pro účely DC KNAV jsou do programu dokumenty vkládány opět ve formátu TIFF zkomprimované LZW kompresí. Sirius sice není schopen pracovat s komprimovanými obrázky, ale sám si je dekomprimuje a dále s nimi pracuje v této podobě. Pro načítání do programu jsou komprimované soubory výhodnější, protože je program rychleji načte než soubory bez komprese.

Dokumenty v hlavním okně Siria jsou organizovány do pořadačů, ve kterých jsou již zpracované dokumenty. Pořadače jsou stromově uspořádány do čtyřech úrovní: hlavní úroveň je Periodikum, jako typ pořadače, druhá úroveň pořadače je titul periodika, třetí úroveň představuje jeden ročník periodika a poslední, čtvrtá úroveň je rezervována pro jednotlivé výtisky periodik. V prostřední části hlavního okna jsou

zobrazeny jednotlivé záznamy dokumentů v označeném pořadači daného výtisku, které si lze zobrazit v pravé části hlavního okna. Metadata neboli indexy až na úroveň jednotlivých stran lze upravovat ještě i po zpracování dokumentů v hlavním okně v levé části pod pořadači.

Mezi indexy a pořadači jsou ještě další části, jako například Rozpracované zakázky (na obrázku jsou skryty), ve kterých jsou rozpracované dávky dokumentů. Dávka může být pod jednou zakázkou jedna, nebo více, podle volby uživatele. V DC KNAV je obvykle vedena jako jedna zakázka jeden ročník daného titulu a k té se přidávají výtisky daného ročníku jako jednotlivé dávky dokumentů. Jednou dávkou dokumentů může být také celý jeden ročník periodika, který se pak rozdělí na jednotlivé výtisky během indexování.

V hlavním okně Siria je k dispozici mnoho různých nástrojů pro nastavení funkcí programu a mohou zde probíhat dodatečné úpravy již zpracovaných dokumentů a to jak po stránce editace metadat, tak po stránce grafické.

Prvním krokem při zpracování nového titulu v Siriovi je nutné založit nový pořadač periodika s názvem titulu a ostatními metadaty vztahujícími se k tomuto titulu. Ještě nedávno bylo nutné ručně vyplňovat metadatová políčka dle katalogu. Dnes je již k dispozici nástroj pro automatické stahování informací o titulech z různých knihoven. Do Siria byl implementován klient Z39.50, který umožňuje okamžitě vyhledat a stáhnout metadata z knihovních systémů. Ta se zapisují do indexových položek Siria na úrovni titulu. Systém indexů v Siriovi byl vytvořen podle DTD pro periodika a monografie definovaném Národní knihovnou.

Metadata jsou stahována z Národní knihovny ze Souborného katalogu ČR vzhledem k předpokladu úplnosti informací. Po stažení informací je nutné indexy překontrolovat a některé informace pozměnit, např. místo uložení dokumentu a jeho signatury, protože dokumenty určené k digitalizaci jsou vybírány z fondů KNAV.

Po vytvoření pořadače s názvem a informacemi daného titulu je nutné nahrát dokumenty do programu. Po vytvoření zakázky následuje vznik dávky, do které se v okně **Dávkové skenování** načtou hromadně dokumenty. V této fázi lze navolit různé možnosti grafického zpracování dokumentů, např. ořezy, vyčištění nebo vyrovnání.

V případě DC KNAV je využívána pouze funkce pro určení pravých a levých stránek v dávce dokumentů. Při následném ukládání celé dávky dokumentů je operátorem zvoleno, které stránky jsou levé a které pravé. Rozeznání pravých a levých stran je základem pro další kroky zpracování. V této fázi se provádí také analýza ořezů, pokud je však povolena v nastavení. Ještě nedávno bylo nutné tímto krokem i samotným ořezem projít, protože bez něj nebylo možné učinit další operace. Protože jsou dokumenty již ořezány z Book Restoreru™, bývala hodnota ořezu nastavena na

nulovou hodnotu. Dokument se v podstatě neořízl, ale prošel nutnou procedurou. Nyní již umožňuje systém tento zbytečný krok vynechat.

Vytváření dávek, jejich ukládání a určení levých a pravých stran probíhá na lokálních pracovních stanicích. Záznamy dokumentů jsou však ukládány do databáze a samotné dokumenty do úložiště na server.

Po uložení dávky dokumentů je otevřeno okno **Kontrola dokumentů**. V současné době již není nutné dokumenty kontrolovat, protože krok ořezání dokumentů je možné vypnout, je však třeba vytvořit šablony pro načítání čísel stránek z obrázků neboli OCR (Optical Character Recognition). Šablona se připravuje pro pravou a levou stranu zvlášť. V každé šabloně se nastaví oblast na obrázku, ze kterého se má číslo strany načíst. Dále se musí určit indexové pole, do kterého se vyčtené informace zapíše. Pokud jde o číslo strany většinou se jedná o index na úrovni stránky periodika a konkrétně položka **Číslo strany**.

Po vytvoření skupiny šablon pro pravou a levou stranu a nastavení funkce **Zónové čtení** je dávka uložena a automaticky zpracována. Pro zónové čtení je nutný hardwarový klíč pro OCR. V Siriovi je implementován program pro OCR FineReader Engine od ukrajinské firmy ABBYY, který tuto funkci vykonává. Hardwarový klíč pro OCR je vždy v jednom počítači, který se v tu chvíli chová jako server. Ostatní uživatelé posílají své dávky na tento server, který jim zónové čtení provede. Tento Hardwarový klíč je přenositelný na jakýkoliv počítač, kde je nainstalovaný Sirius. Serverem tak může být kterákoliv stanice podle potřeby pracoviště. Prakticky je vhodné mít jeden počítač na toto dávkové zpracování vyčleněný, protože pokud zde vykonává Sirius dávkové zpracování nelze v něm zároveň pracovat na jiných činnostech.

Když je dávka dokumentů zpracována, otevře se do okna Indexování. Zde jsou v pravém sloupci zobrazeny indexy ke každé straně, výtisku nebo ročníku, podle toho kde se nachází kurzor v prostředním sloupci. V prostředním sloupci jsou viditelné záznamy jednotlivých stran a jejich obraz v sloupci levém.

V této fázi dochází k vytváření a editaci indexů do indexových polí na úrovni ročníku, výtisku a článků. Indexy z nejvyšší úrovně **Titulu**, které byly staženy na počátku pomocí klienta Z39.50 jsou zobrazeny automaticky. V této fázi je však nelze editovat. V současné době jsou vyplňovány položky indexů k ročníku a výtisku. Od roku 2008 budou popisovány i jednotlivé články.

Indexování má svá speciální pravidla, na jejichž tvorbě se DC KNAV kromě Národní knihovny a firmy Elsys Engineering rovněž podílelo. Tato pravidla vychází částečně z katalogizačních pravidel, knihovnických zvyklostí, ale velká část musela být vytvořena zvlášť pro tento styl tvorby metadat. Pravidla jsou neustále vyvíjena a aktualizována. Pravidla pro popis dokumentů jsou dvojího druhu – pro popis periodik a pro popis monografií. Tato pravidla nejsou nutně závazná pro všechny

digitalizující instituce, ale jsou závazná pro ty, kdo by v budoucnu chtěli spolupracovat na výměně digitalizovaného materiálu s Národní knihovnou ČR.

Pokud jsou všechna potřebná indexová pole vyplněna, dávka je uložena a uvolněna do pořadače v hlavním okně. Indexy lze ještě v hlavním okně dodatečně editovat a doplňovat.

Když jsou obrázky uvolněny do pořadačů, přichází na řadu provedení celostránkového OCR. OCR provádí opět FineReader zaimplementovaný do Siria. Pro OCR je tedy opět nutný hardwarový klíč. V současné době jsou v DC KNAV tyto klíče tři. Dva z nich jsou ve verzi 5.0, ty jsou využívány na vyčítání čísel stran. Jeden je v nové verzi 8.0, který slouží pro celostránkové vyčítání.

Výsledkem OCR jsou pak jednotlivé textové dokumenty s vyčteným textem ke každému obrazu stránky. Tyto vyčtené texty pak slouží pro fulltextové vyhledávání v digitální knihovně. Kromě těchto klíčů pro OCR vlastní DC KNAV ještě samostatnou licenci FineReaderu verze 8.0 přímo od firmy ABBYY. Tento klíč slouží pro samostatné OCR textů, které nemusí procházet všemi procesy úprav, jedná se především o speciální zakázky či testování různých druhů dokumentů.

Z pořadačů lze dokumenty exportovat buď jako celý titul, nebo jenom ročník či výtisk. V DC KNAV je exportován vždy každý ročník z titulu samostatně. Výsledkem exportu je adresář s obrázky ve formátu TIFF a textových dokumentů s vyčteným textem a souborem ve formátu XML, kde jsou popisná metadata k celému titulu, ročníku, výtisku a jednotlivých stran. Ten lze editovat buď ručně v textovém editoru nebo v XML editoru (např. Xmetal).

Zálohování

Zálohování je nedílnou a důležitou součástí digitalizace. Z finančních prostředků projektu DK AV ČR, který byl zmíněn výše, byly pořízeny dvě páskové mechaniky StorageWorks Ultrium 460 od firmy Hewlett Packard. Tyto mechaniky slouží k přenesení dat na pásky. Jedna mechanika je umístěna v digitalizačním centru, druhá je v hlavní budově AV ČR na Národní třídě. Mechanika na digitalizačním pracovišti slouží k zálohování výstupů digitalizace. Konkrétně jsou zálohovány:

- prvotní naskenované obrazy, tzv. originály, přímé výstupy ze skenovacích strojů ve formátu TIFF a LZW kompresi,
- soubory, jenž jsou upravené a připravené ke konverzi do DjVu a importu do Krameria – tj. adresáře s obrázky a vyčtenými textovými dokumenty a xml soubory s popisnými metadaty,
- všechny ostatní naskenované dokumenty v rámci speciálních objednávek, v podobně originálních skenů v LZW kompresi a v podobě upravených obrazů (v případě jejich tvorby).

Druhou páskovou mechaniku využívá oddělení informačních technologií. Využívá ji k zálohování systému Kramerius, zálohovány jsou všechny obrázky v něm uložené a jejich metadata a databáze záznamu obrázků.

Pro ukládání se využívají pásy o kapacitě 400 MB dat. V současnosti je v DC KNAV již 32 plných pásek dat. Podle výrobce je životnost dat na páskách cca 30 let. S obměnou za novější technologii je počítáno přibližně jednou za 4 roky.

ZÁVĚR

Od vzniku DC KNAV uplynulo již 4,5 roku. Začátky nebyly snadné, nikdo z pracovníků neměl zkušenosti s digitalizací. Vycházeli jsme ze získaných poznatků z jiných institucí a především vlastních pokusů a předchozích zkušeností z oboru informačních technologií.

Během těchto let se pracoviště rozvinulo v plně fungující centrum s velkou produkcí digitalizovaných dat, které je flexibilní a neustále se rozvíjí dle nových trendů v oblasti informatiky a digitalizace. V získávání cenných odborných znalostí a poznatků velmi napomáhá spolupráce na různých projektech v oblasti digitalizace vyhlášených nejen Ministerstvem kultury České republiky, ale i Evropskou unií.

SEZNAM POUŽITÉ LITERATURY

1. BARTOŠEK, M. Digitální knihovny – teorie a praxe. Národní knihovna [online]. 2004 [cit. 2007-11-01], roč. 15, č. 4. Dostupný z WWW: <<http://knihovna.nkp.cz/NKKRo404/0404233.html>>.
2. BookRestorer™. In Book Restorer™ : User Manual. I2S, [200-?]. 80 s.
3. BookRestorer™. In The Restore Modules Book Restorer™ : User Manual. I2S, [200-?]. 60 s.
4. BRATKOVÁ, Eva. Dublin Core. In Metadata jako nová nástroj pro komunikaci webovských informačních zdrojů. Národní knihovna [online]. 1999 [cit. 2007-11-14], roč. 10, č. 4, s. 178-195. Dostupný z <<http://full.nkp.cz/nkkkr/Nkkkr9904/9904178.html>>.
5. BRATKOVÁ, Eva. ISSN. In Síť identifikátorů informačních entit: pracovní text pro potřeby výuky posluchačů ÚISK FF UK [elektronický text]. Verze 2.4, Praha: ÚISK FF UK, 20.2.2006. Nепublikováno.

6. BRATKOVÁ, Eva. Metadata. In Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. Národní knihovna [online]. 1999 [cit. 2007-09-12], roč. 10, č. 4, s. 178-195. Dostupný z <<http://full.nkp.cz/nkkr/Nkkr9904/9904178.html>>.
7. CELBOVÁ, Ludmila. Metadata. In KTD : Česká terminologická databáze knihovnictví a informační vědy (TDKIV) [online databáze]. Praha : Národní knihovna České republiky, 2003- [cit. 2007-10-06]. Systémové číslo: 000000543. Dostupný z WWW: <<http://sigma.nkp.cz/>>.
8. DigiBook 2000 LC. In Suprascan range and DigiBook 2000LC, Digitization terminal for registers and books. Pessac Cedex, France: I2S, May 2005.
9. Digitalisieren – internationale Projekte in Bibliotheken und Archiven. Berlin: BibSpider, 2007. ISBN 978-3-936960-17-4.
10. EDL Project. In About EDLproject. [cit. 2007-11-25] Dostupné z WWW: <<http://www.edlproject.eu/about.php>>
11. Formáty papíru. In Wikipedia : Otevřená encyklopedie [online]. St. Petersburg (Florida) : Wikimedia Foundation Inc., Czech version 2002- [cit. 2007-10-30]. Tato stránka byla naposledy editována 23.10.2007. Dostupný z WWW: <http://cs.wikipedia.org/wiki/Form%C3%A1t_pap%C3%ADru>.
12. GNU GPL. In Wikipedia : Otevřená encyklopedie [online]. St. Petersburg (Florida) : Wikimedia Foundation Inc., Česká verze 2002- [cit. 2007-11-15] . Tato stránka naposledy editována 13.11.2007. Dostupný z WWW: <<http://cs.wikipedia.org/wiki/GNU>>.
13. i2010: Digital Libraries. In i2010: Digital Libraries. Luxemburg: Office for Official Publications of the European Communities, 2006 [cit. 2007-11-26]. Dostupné z WWW: <http://ec.europa.eu/information_society/activities/digital_libraries/doc/brochures/dl_brochure_2006.pdf>. ISBN 92-79-02332-2.
14. KNOLL, Adolf. Konference Nové technologie a standardy: digitalizace národního dědictví 2004. In Ikaros [online].2004 [cit. 2007-11-20], roč. 8, č. 7. Dostupný z WWW: <<http://www.ikaros.cz/node/1727>>. URN/NBN:cz-ik1727. ISSN 1212-5075.

15. LHOTÁK, Martin. DML-CZ: Česká digitální matematická knihovna. In Konference Archivy, knihovny, muzea v digitálním světě 2006. Praha : SKIP, 2006 [cit. 2007-11-08]. [Konference Archivy, knihovny, muzea v digitálním světě 2006, Praha, 05.12.2006–06.12.2006, CZ]. Dostupný z:
<<http://skip.nkp.cz/KeStazeni/Archivyo6/Lhotak.ppt>>.
16. LHOTÁK, Martin. Digitalizační pracoviště KNAV a projekt Digitální knihovna AV ČR. In Informace. Č. 2-3 (2005) [cit. 2007-11-08]. Dostupný z WWW:
<<http://www.lib.cas.cz/informace/Informace-2-3-2005.pdf>>. ISSN 1210-8502.
17. LHOTÁK, Martin. Digitální věk v knihovnách. In Služby knihoven – nové trendy. CASLIN 2005. Lednice (CZ), 19.06.2005–23.06.2005 Brno: Moravská zemská knihovna v Brně, 2005 [cit. 2007-11-08]. Dostupný z WWW:
<<http://www.mzk.cz/CASLIN05/dokumenty/sbornik.doc>>. ISBN 80-7051-164-8.
18. LHOTÁK, Martin. Možnosti spolupráce při vytváření digitálních sbírek. In Archivy, knihovny, muzea v digitálním světě 2004. Praha: Národní technické muzeum, 2005 [cit. 2007-11-11]. Dostupný z WWW: <<http://skip.nkp.cz/akcArcho4.htm>>.
19. LHOTÁK, Martin. Open source pro Digitální knihovnu. In Automatizace knihovnických procesů – 11.: sborník z 11. ročníku semináře pořádaného ve dnech 16.–17. května 2007 v Liberci. Praha: ČVUT, 2007 [cit. 2007-09-10]. Dostupný z WWW: <<http://www.akvs.cz/akp-2007/09-lhotak.pdf>>. ISBN 978-80-01-03691-4.
20. METS. In Metadata encoding & Transmission Standard [online]. The Library of Congress. This page was last modified 7 November 2007 [cit. 2007-11-12]. Dostupný z WWW: <<http://www.loc.gov/standards/mets/>>.
21. Minerva. In About Minerva. Minerva Project 2003-04. Last revision 7.1.2004 [cit. 2007-11-26]. Dostupný z WWW: <<http://www.minervaeurope.org/whatis.htm>>.
22. Minerva Plus. In MINERVA plus. Minerva Project 2004/2005. Last revision 21.12.2004 [cit. 2007-11-26]. Dostupný z WWW:
<<http://www.minervaeurope.org/whatis/minervaplus.htm>>.
23. Minerva eC. In MINERVA eC. Minerva Project 2006-11. Last revision 13.3.2007 [cit. 2007-11-26]. Dostupný z WWW:
<<http://www.minervaeurope.org/about/minervaec.htm#>>.
24. Open Archives Initiative. OAI-PMH. In Open Archives Initiative [online]. 2000 - [cit. 2007-11-10]. Dostupný z WWW: <<http://www.openarchives.org>>.
25. SEEDI. In South – Eastern European Digitization Initiative [online] [cit. 2007-11-25]. Dostupný z WWW: <<http://www.ncd.matf.bg.ac.yu/seedi/index.html>>.

26. Sirius. In Sirius, Příručka administrátora. Vyškov: Elsys Engineering, [200 - ?].
27. Sirius. In Sirius, Postup při zpracování dokumentů. Vyškov: Elsys Engineering, [200 - ?].
28. Sirius. In Sirius, Uživatelská příručka. Vyškov: Elsys Engineering, [200 - ?].
29. Skener Zeutchel Omniscan 7000. In Omniscan 7000 User's Manual. Tübingen – Hirschau: Zeutchel GmbH, [200 - ?].
30. Skener Zeutchel Omniscan 7000. In Omniscan 10.0 Software Instalation Rev. 01. Tübingen – Hirschau: Zeutchel GmbH, [200 - ?].
31. Skener Zeutchel Omniscan 7000. In Omniscan 10.X User Manual. Tübingen – Hirschau: Zeutchel GmbH, [200 - ?].
32. ŽÁRA, Jiří; BENEŠ, Bedřich; FELKEL, Petr. Moderní počítačová grafika. Praha: Computer Press, 1998. ISBN 80-7226-049-9.

Jarmila Burgetová
Martin Lhoták
**ČESKÁ DIGITÁLNÍ
MATEMATICKÁ
KNIHOVNA: DML-CZ**

Česká digitální matematická knihovna (DML-CZ) je plnotextová digitální knihovna, která si klade za úkol zpřístupnit českou matematickou literaturu, především matematické časopisy, vybrané monografie a sborníky. Vytváří se z iniciativy Matematického ústavu Akademie věd České republiky na základě projektu 1ET200190513 AV ČR v letech 2005–2009.

Dílčí cíle projektu jsou:

1. akvizice materiálů vhodných k digitalizaci s ohledem na copyright a IPR¹,
2. digitalizace a opatření materiálů metadaty,
3. vytvoření strukturovaných digitálních dokumentů,
4. vybudování digitální knihovny,
5. začlenění DML-CZ do světové digitální matematické knihovny (WDLM).

Matematický ústav AV ČR je koordinátorem projektu. Spoluřešiteli projektu jsou: Ústav výpočetní techniky a Fakulta informatiky Masarykovy univerzity v Brně, Matematicko-fyzikální fakulta Univerzity Karlovy v Praze a Knihovna Akademie věd AV ČR. V současné době je zpřístupněno více než 7 000 matematických článků, představujících přes 90 000 stran textu.

Přestože bude řešení projektu pokračovat ještě v roce 2009, mohou již nyní uživatelé přistupovat do zveřejněné beta verze na adrese <http://dml.cz/>. Na těchto stránkách jsou také podrobnější informace o projektu a jeho řešitelích.

Oficiální představení projektu České matematické digitální knihovny (DML-CZ) veřejnosti proběhlo dne 11. června 2008 na tiskové konferenci a mělo v médiích a mezi odbornou veřejností značný ohlas.

Z obsáhlé zprávy ČTK vyjímáme:

„Digitalizace matematické knihovny čile pokračuje a má se stát součástí evropského celku. Čeští vědci dnes zpřístupnili na internetu pilotní projekt České digitální matematické

¹ Intellectual property rights = intelektuální (duševní) vlastnická práva

knihovny (DML-CZ). Matematika je podle nich věda, která nestárne, a pokud matematik neudělá chybu, pak výsledky jeho práce platí ,na věčné časy'. Vědci proto citují i velmi staré práce, mnohdy i starší sta let. Navíc matematické literatury rychle přibývá. Digitalizace matematické literatury tedy byla nutná, řekl na dnešní tiskové konferenci duchovní otec projektu Jiří Rákosník z Matematického ústavu Akademie věd.

Projekt řešitelé plní už od roku 2005 a pokračovat budou do roku 2009. Přejde podle Jiřího Rákosníka zhruba na osm milionů korun, což je na rozdíl od fyzikálních či astronomických projektů zanedbatelná suma. O matematice se říká, že je to levná věda. ,Je to klišé, které nám škodí, protože si všichni myslí, že potřebujeme jen tužku a papír,‘ dodal. Své síly při vytváření digitální knihovny spojili vědci z pěti pracovišť. Kromě Matematického ústavu AV na něm spolupracují odborníci z Ústavu výpočetní techniky a Fakulty informatiky Masarykovy univerzity v Brně, Matematicko-fyzikální fakulty Univerzity Karlovy a z Knihovny AV. Vysoké školy do projektu zapojily také studenty.

Nyní už má tým naskenováno zhruba 100 000 stran matematické literatury. Celkem ale má být do digitální knihovny zařazen dvojnásobek stran. K dispozici by měla být veřejnosti zdarma bez přístupových hesel. Podle Martina Lhotáka z Knihovny AV je ambicí týmu zapojení DML-CZ do budoucí Evropské digitální matematické knihovny (EuDML).

Podle viceprezidenta Evropské matematické společnosti Pavla Exnera je projekt EuDML podán. Cílem je propojit existující i nově vznikající národní digitální archivy matematické literatury a umožnit tak uchování těchto dokumentů, jejich provázání, internetový přístup k nim a vícejazyčné vyhledávání. EuDML by se měla stát součástí budované univerzální celosvětové digitální knihovny.

Projekt musí obstát ve veřejné soutěži o peníze, které EU dává na vědu. Je už v pořadí třetí. Autoři jsou však přesvědčeni, že tentokrát uspěje. Jeho rozpočet je kolem dvou milionů eur (asi 50 milionů korun). V Česku nejde o první digitalizaci literatury. Podle Martina Lhotáka zpřístupnila například Národní knihovna dva miliony stran manuskriptů a starých tisků a tři a půl milionu stran periodik z 18. a 19. století. Také Knihovna AV má svůj projekt. Kvalita digitalizace matematické literatury je podle něj na nejvyšší úrovni. Jiří Rákosník uvedl, že jeho tým je připraven dát svou koncepci k dispozici nejen tuzemským pracovištím, ale i Evropě.“

11.–12. června 2008 se na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze uskutečnil mezinárodní seminář „Launching the DML-CZ. Czech Digital Mathematics Library“ věnovaný odborným otázkám digitalizace matematických textů a digitálních knihoven. Setkali se zde pozvaní odborníci a zástupci digitálních matematických knihoven z několika evropských zemí. Nechyběl ani prezident Evropské matematické společnosti Ari Laptev ze Švédska. Na semináři bylo předneseno 14 příspěvků českých i zahraničních odborníků:

- From pixels and minds to the mathematical knowledge in digital library (J. Rákosník, P. Sojka)
- Scanning and adjusting the images (M. Lhoták)
- Metadata and their sources (J. Veselý, O. Ulrych)
- Metadata Editor (M. Bartošek)
- The digital library DML-CZ (V. Krejčíř)
- Similarity of mathematical texts (R. Řehůřek)
- An automated processing of journal articles for a digital library (P. Sojka)
- A vision for the near-future Digital Mathematics Library (T. Bouche)
- About the RusDML (B. Wegner)
- Digitization projects in Spain (E. Macías-Virgós)
- Springer digital library and related activities (O. Chiarcos)
- Towards the best repository – migration of the Polish Virtual Library to Yadda platform (K. Zamlynska)
- The DML and digital preservation – interoperability over time for mathematical resources. (M. Dobrevá)
- Integrating OAI metadata: Experiences and challenges (T. Fischer)

Na stránkách věnovaných projektu <http://project.dml.cz>, jsou v sekci Documents přístupné příspěvky přednesené na tomto semináři. Kromě toho jsou ve stejné sekci i příspěvky řešitelů přednesené na předchozích tuzemských i zahraničních odborných akcích. a k nahlédnutí je také návrh projektu podaný v roce 2004.

O České matematické digitální knihovně (DML-CZ) odbornou veřejnost přehledně informoval na konferenci INFORUM 2008 RNDr. Miroslav Bartošek, CSc., vedoucí Knihovnicko-informačního centra Ústavu výpočetní techniky Masarykovy univerzity v Brně. Ve svém vystoupení, které je přístupné na internetu (<http://www.inforum.cz/sbornik/81/>), se zaměřil na tři aspekty DML-CZ. Prvním bylo představení samotné digitální knihovny, jejího obsahu a služeb, druhým pak podrobný popis tvorby DML-CZ prostřednictvím Metadatového editoru – nástroje pro vytváření struktury a metadat digitalizovaných objektů (strukturálních, administrativních a popisných metadat, souborů referencí, autoritních záznamů aj.), který byl vyvinut

v průběhu řešení projektu. Třetím aspektem byla charakteristika postupu při začleňování born-digital² materiálu – automatizované přebírání aktuálních časopiseckých čísel.

Akademickou veřejnost seznámila s projektem DML-CZ ve svém článku „Matematická literatura nestárne“ Mgr. Marina Hužvárová, šéfredaktorka Akademického bulletinu.³

Projekt České digitální matematické knihovny je koncipován tak, aby byl životaschopný i po ukončení financování na konci roku 2009. Server by měl být provozován v Matematickém ústavu AV ČR a procesy příjmu dalších dokumentů zejména born-digital (aktuální články časopisů) budou realizovány s důrazem na maximální automatizaci.

² „Born Digital“ – Označení pro digitální materiály, které byly vytvořeny přímo v elektronické formě. Výraz je používán zejména pro odlišení od digitálních materiálů, které byly do elektronické formy převedeny z původní analogové (např. skenováním tištěných knih).

³ Akademický bulletin, 2008, č. 7-8, s. 18-19.

PERSONALIA

INFORMACE by nebyly úplné, kdybychom svým kolegyním a kolegům nesdělili některé osobní zprávy:

PhDr. Ivaně Kadlecové – od prosince 2007 Laiblové Kadlecové – předal předseda Akademie věd prof. Václav Pačes čestnou medaili „Za zásluhy o Akademii věd České republiky“ při neformální slavnosti dne 24. ledna 2008. Dr. Kadlecové byla medaile udělena na doporučení Akademické rady za její úspěšné působení ve funkci ředitelky KNAV od roku 1992 do června 2007. Čestnou medailí oceňuje vedení Akademie věd mimořádné aktivity v oblasti organizování vědy a zajišťování infrastruktury výzkumu a vývoje.

Čestnou medailí Vojtěcha Náprstka za popularizaci vědy obdržel dlouholetý pracovník (a poradce bývalé ředitelky) KNAV Mgr. Antonín Vítek, CSc. Předání medaile se uskutečnilo na slavnostním ceremoniálu dne 29. ledna 2008 ve vile Lanna.

Dr. Vítkovi od jara 2008 na nebi svítí planetka objevená v dubnu 2000 v Ondřejově českými astronomy P. Kusnirákem a P. Pravcem. Planetka má číslo 30253 a nese jméno Vítek.

V roce 2008 zakončily své vysokoškolská studia úspěšnou obhajobou diplomových prací tři pracovnice KNAV:

Mgr. Anna Ryšánková,
Mgr. Iva Burešová

a Mgr. Monika Friedmannová.

Až na poslední jmenovanou, která z KNAV odešla, uvedené kolegyně nás v INFORMACÍCH seznámí s vybranými pasážemi ze svých diplomek.

Již v tomto čísle tak učinila

Mgr. Anna Ryšánková.

Oběma vyznamenaným k udělení čestných medailí a oběma kolegyním k úspěšnému zakončení vysokoškolských studií upřímně
BLAHOPŘEJEME!

K 1. lednu 2008 odešly do důchodu dlouholeté pracovnice revizního oddělení Knihovny AV ČR Marcela Dedíková a Dagmar Blažejová. Navždy nás předčasně opustila pracovitá a vždy ochotná pracovnice knihárny paní Božena Odehnalová.

OBSAH

Úvodem (J. Burgetová)	3
Digitalizační centrum Knihovny AV ČR (A. Ryšánková)	5
Česká digitální matematická knihovna: DML-CZ (J. Burgetová–M. Lhoták)	33
Personálie	37
Obsah	38



PRO PRACOVNÍKY STŘEDISEK VĚDECKÝCH INFORMACÍ
A KNIHOVEN AV ČR VYDÁVÁ KNIHOVNA AV ČR, v.v.i.

REDIGUJE PhDr. JARMILA BURGETOVÁ

SRPEN 2008

ISSN 1210-8502