



Dňa 11. mája 2010 podpredseda vlády SR a minister školstva SR Ján Mikolaj slávnostne otvoril v Centre vedecko-technických informácií SR (CVTI SR) pracovisko na digitalizáciu obsahu kníh, tlačovín a dokumentov pre slovenskú vedu. Toto pracovisko bolo vybudované vďaka finančnej podpore Európskej únie, konkrétne Európskeho fondu regionálneho rozvoja v rámci realizácie národného projektu „Infraštruktúra pre výskum a vývoj – Dátové centrum pre výskum a vývoj“.

Ide o moderné pracovisko na digitalizáciu a následné spracovanie a sprístupnenie dokumentov pre vedeckú komunitu, ktoré umožňuje efektívne uspokojiť požiadavky na digitalizáciu objektov rôzneho typu, či už sú to knihy, časopisy alebo napríklad mapy nachádzajúce sa vo fondoch Slovenskej akadémie vied. V digitalizačnom pracovisku sa nachádzajú skenery s dostatočným výkonom na digitalizáciu nasledujúcich typov dokumentov:

- v papierovej podobe pomocou automatického podávača (všetky dokumenty, ktoré je možné rozviazať na voľné listy),
- zviazané dokumenty na plochých lôžkach (napr. dokumenty zviazané a pečatené, dokumenty s hrúbkou niekoľkých mm a pod.),
- veľkoformátové dokumenty (veľkoplošné dokumenty do rozmerov A0),
- knihy (dokumenty zviazané knižnou väzbou),
- mikrofily (dokumenty už nasnímané na mikrofily).

Zoskenované dokumenty je možné previesť konverziou OCR s možnosťou kontroly kvality do editovateľných formátov a ďalej upravovať a systematicky ukladať s možnosťou plnotextového vyhľadávania. Toto pracovisko je unikátne najmä pre svoju univerzálnosť, ale aj výkonom sa radí na Slovensku k významným pracoviskám tohto druhu. Postupne sa uvádza do prevádzky viacero špičkových zariadení, z nich je finančne najnáročnejší aj výkonom najpozoruhodnejší poloautomatický knižný skener (inak nazývaný aj knižný robot), ktorý umožňuje rýchle a pohodlné spracovanie kníh bez nutnosti porušenia ich väzby. O výkone zariadenia si možno urobiť predstavu podľa toho, že oskenovanie osemstostránkovej knihy nezaberie viac ako 60 minút.

Okrem tohto robustného zariadenia zo Švajčiarska (váži vyše jednej tony) má digitalizačné pracovisko k dispozícii mikrofilmový skener, dva veľkoplošné skenery typu plotter (akceptuje dokumenty do šírky 2A0), 2 výkonné skenery na jednotlivé listy a 3 produkčné skenery s plochým lôžkom. Digitalizačné pracovisko v CVTI SR v Bratislave bude digitalizovať obsah a ďalej s ním pracovať tak, aby sa dosiahol konečný cieľ – často unikátny obsah uložiť, archivovať, no najmä sprístupniť vedeckej komunite. Pri týchto činnostiach sa počíta najmä s využitím kapacity a možností Dátového centra pre výskum a vývoj na Žilinskej univerzite v Žiline (viac o projekte Dátové centrum pre výskum a vývoj v rubrike PROJEKTY na strane 31) budovaného v rámci tohto istého projektu. Pripravené výstupy z digitalizačného pracoviska sa po sieti dostanú prakticky okamžite do Dátového centra pre výskum

a vývoj, kde sa budú archívovať a sprístupňovať. Plánované kapacity a výkon centra vytvárajú predpoklady na to, aby produkcia digitalizačného pracoviska (inštalované zariadenia sú schopné vygenerovať za jediný pracovný deň až 1 TB digitálneho obsahu) bola jednak bezpečne uložená a jednak sprístupnená odbornej verejnosti.

Proces digitalizácie

Proces digitalizácie sa skladá z nasledujúcich krokov:

- Príjem a evidencia zapožičaných originálnych dokumentov na digitalizáciu.
- Skenovanie strán primárneho dokumentu podľa požiadaviek.
- Úprava digitálneho obrazu (nastavenie kontrastu a jasú, odstránenie šumu, natočenie, orezanie čiernych okrajov a pod.).
- Výstupná kontrola kvality digitálnej kópie podľa určeného pracovného postupu. V prípade nekvality opätovná digitalizácia.

Digitalizačné pracovisko nie je dodnes komplexne dobudované. Momentálne sa pracuje na koncepcii digitalizácie a v rámci projektu Dátové centrum pre výskum a vývoj (DC VaV) sa obstaráva softvér na optimalizáciu výkonu skenovacej linky v oblasti hromadnej digitalizácie.

Typy skenerov

Digitalizačné pracovisko používa šesť typov skenerov. Najzaujímavejší je skenovací robot DL 3003 od Švajčiarskeho výrobcu 4 digital books. Robot o rozmeroch približne 3 x 1,5 x 2 metra dokáže samostatne skenovať zviazané knihy bez minimálnych zásahov operátora. Automaticky detekuje parametre knihy a jeho flexibilný algoritmus dokáže nájsť najlepší spôsob, ako otáčať strany počas skenovania. O skeny sa starajú dve kamery, ktoré dokážu zachytiť obraz v rozlíšení 200, 300, 400 a 600 DPI od formátov A5 až po A2. Rýchlosť skenovania je možné nastaviť, takže je zabezpečené skenovanie ako nových kníh s kvalitnou väzbou, tak aj starších krehkých typov s jemným papierom a voľnou väzbou.



O skenovanie voľných listov sa starajú dva typy skenerov. Jedným z nich je Canon DRX-10C (digitalizačné pracovisko má dva kusy). Tento prístroj dokáže obojstranne oskenovať približne 250 listov za minútu (500 skenov). Taktiež zvláda mnoho funkcií, ako je potlačanie presvitu na skenoch alebo „dokreslenie“ chýbajúcich častí, ako sú napríklad odtrhnuté rohy strán a podobne. Rozlíšenie skenov je maximálne 600 DPI.



Podobný typ, ako je Canon DRX-10C, je aj produkčný skener Kodak i1440 (tri kusy), ktorý tiež skenuje voľné listy. Rýchlosť produkovaného skenovania je pomalšia, zato lepšie zvláda fotografie a produkuje skeny v maximálnom rozlíšení 1 200 DPI. Súčasťou tohto skenera je aj ploché ložko na skenovanie formátov A3.



Mikrofilmový skener MS 800 II skenuje mikrofilmy a mikrofiše. Zvládne skenovať čiernobiele obrázky a obrázky v odtieňoch sivej farby do rozlíšenia 600 DPI a veľkosti A3.



Dva veľkoplošné skenery Océ CS4354QS sú pomocníkom pri skenovaní veľkorozmerných formátov, hlavne máp, a dokážu skenovať do rozlíšenia 1 200 DPI a formátu 2A0.

Skener Minolta PS 5000C MKII je zakúpená z vlastných zdrojov CVTI SR. Je to skener na ručné skenovanie, vhodný na skenovanie voľných listov aj zviazaných kníh a časopisov do veľkosti A3 a s maximálnym rozlíšením 600 DPI.



Skenovanie a plány do budúcnosti

Počas skoro dvojročnej prevádzky digitalizačného pracoviska sa skenovalo pre veľa ústavov Slovenskej akadémie vied. Ide napríklad o Jazykovedný ústav Ľudovíta Štúra, Ústredný archív, Geologický ústav, ale aj mnoho ďalších. Doteraz bolo zdigitalizovaných približne dva a pol milióna strán pre ústavy SAV, ako aj pre interné potreby CVTI SR. Digitalizačné pracovisko naďalej kontaktuje ústavy a propaguje možnosť digitalizovania dôležitých, prípadne vzácných materiálov, a digitalizuje obsahy prírastkov knižničného fondu CVTI SR. V blízkej budúcnosti sa plánuje zakúpenie workflow softvéru a prezentačnej vrstvy, kde by sa materiály mohli zdieľať. Workflow by optimalizoval a efektívne riadil poradie a dôležitosť prác a tak isto

by organizoval a priradoval prácu konkrétnym skenerom či pracovníkom DP. Softvér prezentačnej vrstvy by umožnil ústavom SAV prezentovať svoje materiály bez nutnosti nahrávania na externé pevné disky alebo USB kľúče. CVTI SR by týmto spôsobom mohlo prezentovať prírastky knižnice cez webové prostredie, čím by užívatelia mohli vyhľadávať informácie z pohodlia domova.

Skenovanie pre verejnosť

Pracovníci digitalizačného pracoviska radi privítajú záujemcov o skenovanie materiálov. Pri práci sa riadia autorským zákonom, ktorý dovoľuje kopírovať nepodstatnú časť publikácie, a platným cenníkom stanoveným Centrom VTI SR.

Ing. Mikuláš Kevély
mikulas.kevely@cvtisr.sk
odbor informačných technológií

Bc. Pavol Marcina
pavol.marcina@cvtisr.sk
odbor informačných technológií