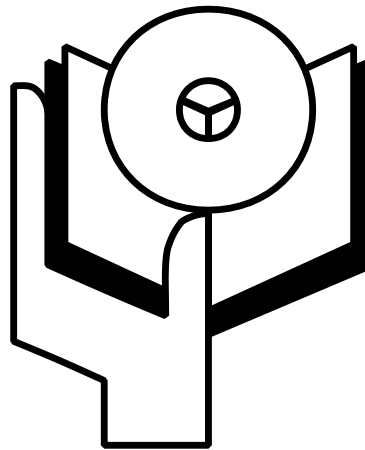


lossiger- Rod- Katuščák

Metodológia digitalizácie písomného dedičstva



4 DigitalBooks™

Skenery 4DigitalBooks

- Automatické knižné skenery

Digitizing Line

Exkluzívna technológia založená na automatickom bezobslužnom obrazení strán

- Manuálne knižné skenery

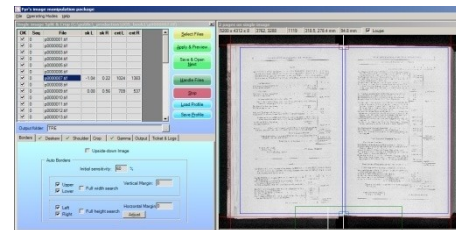
CopiBook & Suprascan

i2S spoločnosť je podielnikom a obchodným partnerom

- Softvér na zlepšovanie obrazu

Page Improver

Exkluzívna technológia automatického delenia a usporiadania strán



Významné znaky Digitizing Line

- Integruje špičkovú kvalitu skenerov i2S a ponúka obraz vo veľkosti od **70 to 112 Mega Pixelov optickej kvality**
- Prispôsobené pre staré i súčasné knihy, certifikované expertami z oblasti ochrany kníh
- Spracúva knihy s rôznou hrúbkou papiera, úrou a pórovitosťou



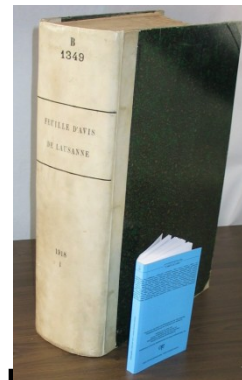
Významné znaky Digitizing Line

- Výnimočne kvalitný obraz celého povrchu dokumentu
- Dokument sa jemne pritláča ku sklu, aby sa zabezpečilo rovnaké celej plochy strany
- Použitím skla sa predchádza zalomeniu strán v chrbtovom zlome knihy
- Automatické zastavenie pri detekcii chyby (spojené strany, nerozrezané listy)



Významné znaky Digitizing Line

- **Pokrýva 98% potrieb knižnice**
 - Spracuje formáty od A6 do A2
 - Spracuje ľahkú (brož) aj tvrdú väzbu
 - Rýchlosť skenovania – do 3000 strán za hodinu/1 robot
 - Spracuje knihy, ktoré sa spontánne zatvárajú
 - 30% moderných kníh, ktoré sa spontánne zatvárajú
 - 50% formátov A6/A5 kníh, ktoré sa spontánne zatvárajú
- Digitizing Line je jediný stroj, ktorý spracuje všetky uvedené kategórie kníh



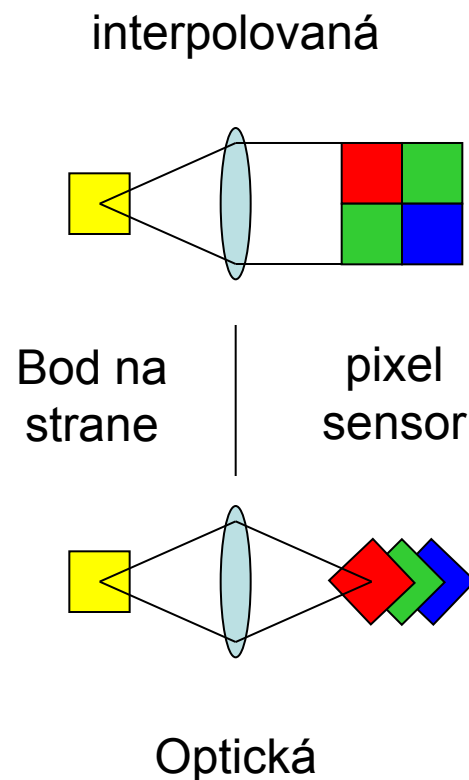
Funkcionalita kamery

Optická = Nie interpolovaná

– Plné optické rozlíšenie

nejde o interpolovanú Bayer
matricu používanú v
digitálnych kamerách

– Pre každý bod zo strany
jeden Pixel Sensor v
kamere



Funkcionalita kamery

- Použité kamery nemajú mechanickú spúšť
 - Bežné mechanické spúšte komerčných kamier končia životnosť pri cykle 400 000
 - Robot dosiahne túto hranicu za 3 mesiace





Partnerstvo excelentnosti

Skenovacia technológia

DigiBook
CopiBook



OCR technológia

ABBYY – DEFINING RECOGNITION



ABBYY® Recognition Server

**Server-based OCR and
PDF Conversion Solution**

Pracovný plán vybudovanie centra digitalizácie

1. Príprava fyzických zdrojov
2. Organizácia toku materiálu
3. Organizácia toku dát
4. Založenie škálovateľnej architektúry
5. Od labáku k fabrike
6. Projekt škálovateľnej architektúry
7. Projekt procesu
8. Príprava organizácie
9. Dimenzovanie výkonnosti digitálnej fabriky

Príprava fyzických zdrojov

1/2

- **Miesto**

- Operačný priestor / budova a miestnosti
- Dispozičný plán pracovísk a zdrojov
- Dostupnosť zdrojov (ergonómia)
- Tok materiálu (kontajnery, knihy)

- **Infraštruktúra**

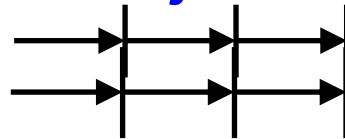
- regulácia svetla
- klimatizácia
- elektrická sieť
- dátová sieť
- bezpečnostný systém

- Zariadenie

- pracovné miesta, stoly, nábytok
- autá, prepravky a kontajnery na knihy
- skenery (ručné, automatické, sheet-feed, sk. mikrofilmov, mikrofilmy zo skenov ...)
- počítače PC (spracovanie obrazu, OCR, indexovanie)
- dátové servery, záložné servery
- repozity/storage, páskové knižnice, disky
- archivačný systém/systemy
- sejfy na uskladnenie vzácných materiálov

Organizácia toku materiálu

- Preberanie kníh na digitalizáciu
- Odovzdávanie kníh po digitalizácii
- Triedenie prepraviek a kontajnerov kníh
- Triedenie kníh / barkód, RFID
- Manažment **paralelnej** dávkovej **produkcie**



- Vyhnúť sa **hromadeniu** a **zápcham**



- Vyhnúť sa **opravám chýb** v hlavnej linke



Organizácia toku materiálu

Digitálna knižnica – digitalizačná technologická linka

System digitálnej knižnice (prvky systému)

- Digitálna akvizícia (príprava)
- Fyzická digitalizácia (technológia masovej a manuálnej digitalizácie)
- Digitálna „katalogizácia“ (popisné metadáta)
- Digitálny archív (dlhodobé archivovanie v CDA, inštitucionálny repozit)
- Digitálne služby (sprístupňovanie digitálneho obsahu)
- Správa prístupov, marketing a copyright (právne, ekonomické, technické otázky sprístupňovania digitálneho obsahu)

Technologické triedy - Zdroje digitálnej akvizície

- **Akvizícia dokumentov (Odbor doplňovania a spracovania dokumentov napr SNK) (ca 8000-10000 titulov ročne)**
- **Depozit napr SNK (ca 500 000 titulov)**
- **Akvizícia a spracovanie seriálov napr SNK (ca 10 000 titulov)**
- **Depozity seriálov u partnerov (časti chýbajúce ročníky, čísla, strany)**
- **Regionálne noviny a časopisy vo verejných knižniciach**

Technologické triedy

- 1. Digitalizované mikrofilmy SNK a partnerov**
- 2. Webharvesting**
- 3. Originálne elektronické zdroje (born digital)**
- 4. Historické knižničné dokumenty (HKD) SNK a partnerov**
- 5. Digitalizácia na požiadanie pre fyzické osoby (prinesené dokumenty)**
- 6. Digitalizácia na požiadanie pre právnické osoby, knižnice, školy, občianske združenia, organizácia, podnikateľské subjekty (prinesené dokumenty)**
- 7. Digitalizácia na požiadanie pre právnické osoby (prinesené/privezené dokumenty)**

Príklad analýzy procesov z jedného zdroja

Odbor doplňovania a spracovania dokumentov SNK

- Odbor kompletne zabezpečuje doplňovanie knižničných fondov SNK a jeho odborné spracovanie podľa schválených medzinárodných a národných štandardov.

Oddelenie doplňovania

- Doplňuje knižničný fond SNK. Zabezpečuje súbežné a retrospektívne doplňovanie slovacikálnych dokumentov (s výnimkou periodík) a výberové doplňovanie zahraničných dokumentov (s výnimkou nákupu a darov periodík).

Procesy

1. *Príjem dodaných dokumentov v akvizícii*
2. *Evidencia dodaných dokumentov v akvizícii*
3. *Pridelenie čiarového kódu/resp RFID jednotlivému exempláru*
4. *Pridelenie jednoznačnej identifikácie každému samostatnému fyzickému objektu*
5. *Nalepenie čiarového kódu/RFID*
6. *Zaznamenanie údajov do RFID*
7. *Príprava dokumentov na vstupné akvizičné spracovanie*
8. *Vyhotovenie skráteného katalogizačného popisu*
9. *Výber slovacík na digitalizáciu zo zdrojov – povinný výtlačok, kúpa, výmena, dar*
10. *Fyzická kontrola stavu exemplára (rozrezanie spojených strán, zhodnotenie formátu, oddelenie samostatných príloh do osobitného obalu s exemplárom,*
11. *Uloženie vybratých slovacík na digitalizáciu do prepravného kontajnera*
12. *Prevoz kontajnera k digitalizačnému robotu*
13. *Vyhotovenie záznamu o odovzdaní a prebratí obsahu kontajnera medzi operátorom akvizície a operátorom digitalizácie*

Katalogizácia

14. *Určenie objektov na katalogizáciu do osobných podadresárov v adresári Katalogizácie (manažér katalogizácie)*
15. *Spracovanie úplného bibliografického záznamu na základe digitálneho objektu exemplára*
16. *Vyhľadanie digitálneho objektu (katalogizátor)*
17. *Otvorenie digitálneho objektu (katalogizátor)*
18. *Vytvorenie bibliografických popisných metaúdajov MARC 21 podľa digitálneho objektu*
19. *Záznam informácie o existencii digitálnej verze popisovaného titulu (MARC 21, tag 530)*
20. *Vloženie jednoznačnej identifikácie digitálneho objektu do záznamu MARC 21*
21. *Vytvorenie holdingového záznamu MARC 21*
22. *Vytvorenie záznamu exemplára MARC 21*
23. *Vloženie linky – spojenie záznamu MARC 21 s digitálnym objektom (MARC 21 tag 856)*
24. *Vloženie URL adresy digitálnej verzie exemplára do záznamu MARC 21*
25. *Uloženie záznamu do databázy ILS*
26. *Synchronizácia ILS Virtua s DIS Vital (Konverzia onfly MARC 21 – XML, kvalifikovaný Dublin Core)???*

Digitalizácia (čierna skrinka)

Digitalizácia (komplexný technologický proces)

Uloženie digitálneho objektu do adresára

Katalogizácia (administrátor digitalizácie)

Organizácia toku dát

1/2

- Manažment tvorby súborov, každá výrobná etapa tvorí nové súbory (TIFF, JPEG, XML, TMP, PDF, ...)
- Manažment pohybu súborov do ďalšieho procesu
- Zálohovanie produkčných dát
- Archivovanie produktov do repozitu
- Riadenie a obnova produkčných pamätí
recycling production storage
- Kontrola výrobného procesu
- Kontrola digitálnych produktov

Organizácia toku dát

2/2

- Nezamieňať funkcie **produkčných pamätí** (obnovuje sa) s **repozitnou pamäťou** (rastie)
- Uprednostňovať distribuované produkčné pamäte pred centrálnymi kvôli vyhnutiu sa konkurenčnému čítaniu/zapisovaniu na serveroch
- Myslite na to, že treba: „**zväčšovať diskový priestor**“

Od labáku k fabrike

L Labák je riadený niekoľkými vysokokvalifikovanými technikmi

F Fabrika je riadená celou organizáciou

L Za mnoho úloh zodpovedá len niekoľko ľudí

F Za niekoľko úloh zodpovedá mnoho ľudí

L Procesy sú len v hlavách niekoľkých ľudí

F Inteligencia je v procesoch a nie len v hlavách ľudí

L Prioritou je manažment technológie a kvality

F Prioritou je manažment zdrojov a kvantity

Profily zamestnancov digitalizačného pracoviska

5 profilov

Profil č. 1

Sekretárska úroveň

- Príjem a registrácia objektov
- Vyhľadanie, zápis metaúdajov
- Príprava a transport vozíkov
- Inšpekcia a kvalifikácia kníh
- Výber príslušnej produkčnej línie
- Používanie počítača, MS Excel
- Odhlásenie, vrátenie, doručenie „zákazníkovi“
- Nie pre dokumentaristu
- **1 pilotná prevádzka / 3-4 továreň**

Profil č. 2

Operátori skenerov

- Základné technické zručnosti
- Schopný zvládať prácu s vysokou úrovňou opakovania, schopnosť a učiť viest' ostatných
- Základy používania počítača, základná manipulácia s objektmi
- Dodržiava pravidlá inšpekcie súborov
- 1 pilotná prevádzka / 3-4 továreň

Profil č. 3

Úprava a manipulácia s obrazmi

- Schopný používať Photoshop, Quark Express
- Grafik, fotograf, polygraf a pod.
- Chápe rozloženie dokumentov
- Pokročilý používateľ PC (správa súborov a adresárov)
- Chápe pixely, farbu, formáty súborov
- **2-4 pilotná prevádzka / 4-8 továreň**

Profil č.4

Štruktúrovanie textu

- Rýchly strojopis
- Triedenie súborov, kapitol, názvov, strán
- Operátor na vstupe (zápise)
- Ovláda softvér
- Akceptuje opakujúcu sa prácu
 - 1. úroveň - titul
 - 2. úroveň – kapitola, vnútorný obsah
- **1 pilotná prevádzka / 3-4 továreň**

Profil č.5

Archivácia, agregácia

- Dokumetarista
- Vysoká úroveň skúseností s počítačmi
- Dobré zručnosti pri organizácii digitálnych údajov
- Pochopenie koncepcie digitálnej archivácie
- Napĺňanie digitálneho úložiska
- 6-9 pilotná prevádzka / 13-19 továreň

Počty ľudí

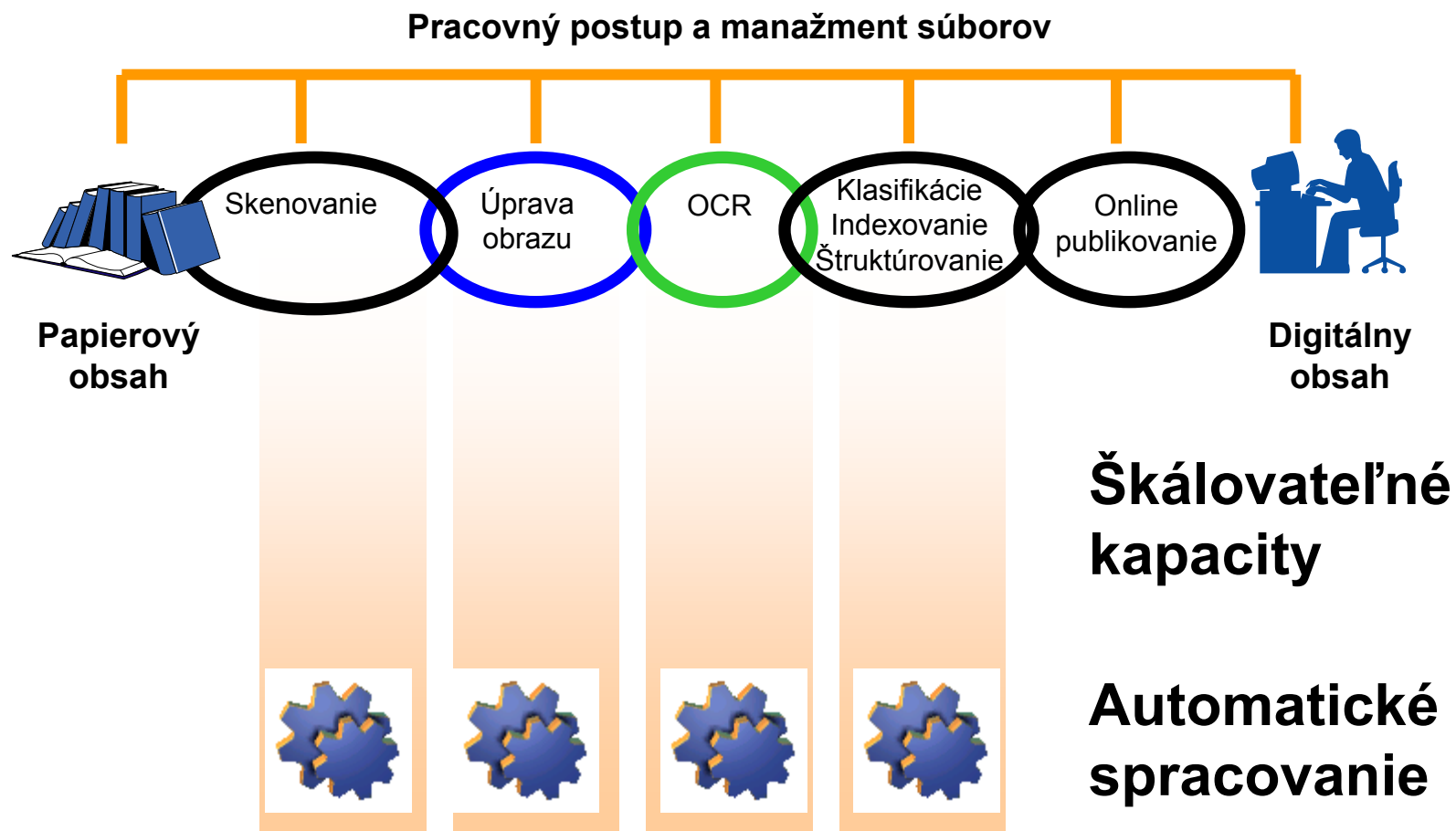
- 1 pilotná prevádzka / 3-4 továreň
- 1 pilotná prevádzka / 3-4 továreň
- 2-4 pilotná prevádzka / 4-8 továreň
- 1 pilotná prevádzka / 3-4 továreň
- 6-9 pilotná prevádzka / 13-19 továreň

11-16 pilotná prevádzka / 26-39 továreň

Projekt škálovateľnej architektúry 1/2

Škálovateľnosť (štruktúra) toku dát je podstatná. Musí byť organizovaná starostlivo, aby sa zabezpečila plynulosť.

Projekt škálovateľnej architektúry digitalizácie v SNK



Projekt procesu

1/2

Príprava organizácie

- Výrobný manažér
 - Stará sa o výrobný personál a zdroje
 - Procesný inžinier
 - Stará sa o výrobné procesy a priepustnosť systému
 - Systémový inžinier
 - Stará sa o IKT
 - Digitalizačný expert
 - Stará sa o dodržiavanie štandardov, implementáciu najlepších skúseností a riešení a kvalitu
 - Kontrola výroby
 - Stará sa o operátorov a výrobu
 - Operátori (pozri nasledujúcu stranu)

Dimenzovanie digitálnej fabriky v SNK Books

Hardvér

automatický knižný skener	1	2	4
ručný knižný skener	1	2	2
pracovné stanice na spracovanie obrazu	2	4	8
OCR servery	1	1	2
pracovné stanice administrátora a na indexovanie	2	2	3

Ľudia

	4	7	12/36
operátori (preberá + pripravuje + skenuje + kontroluje knihy)	1	2	2
operátori (image enhancement + kontrola kvality skenov)	2	4	8
operátori(OCR + indexovanie + odovzdáva + administruje)	1	1	2

Kapacita (strany/pracovná smena)

15'000

30'000

60'000

Veľkosť fabriky

MALÁ

STREDNÁ

VEĽKÁ

Dimenzovanie fabriky - s Digital Books

Príklad: Projekt: 500 000 kníh

= ca 125 000 000 strán

- Produktivita za smenu/prac. deň

= 60 000 strán

- Potrebný čas:

= 2 083 dní

2 083 dní = 9,5 roka...

3 smeny = 3,5 roka

Z toho: 80% masové skenery

20% ručné skenery

Kontakt

Ďakujem za pozornosť

dusan.katuscak@fhv.uniza.sk