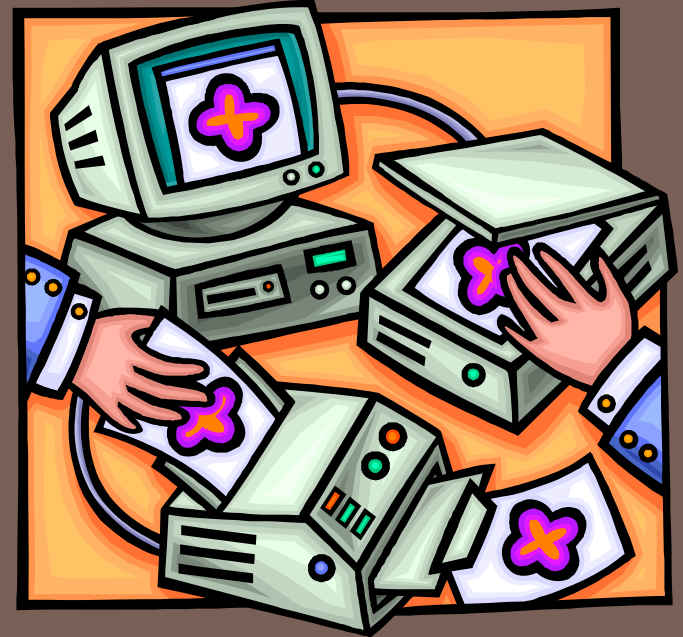


DIGITALIZACE



Kristýna Knapová

Osnova

- Digitální zpracování obrazu
 - ▣ Pixely, RGB model, formáty pro ukládání, komprese
- Postup při digitalizaci
 - ▣ Návrh struktury digitálního archivu
 - ▣ Příprava, skenování, úpravy digitální podoby, pořizování metadat, zařazení do archivu
- Skenovací zařízení
 - ▣ Velkoformátové skenery, skenery s automatickým podavačem formátů A4 a A3
- DMS systémy pro ukládání dat
 - ▣ ELO Professional, EMC Documentum
- Jiné aplikace používané při digitalizaci
 - ▣ ScanWork a další.K

Digitální zpracování

- Pod pojmem digitální zpracování obrazu myslíme veškerou manipulaci s obrazy v digitální podobě.
- Digitální obraz je obraz, který se skládá z velkého počtu obrazových bodů, tzv. **pixelů**, uspořádaných v řádcích a sloupcích.
- Každý obrazový bod má svoji barvu, kterou lze popsat pomocí čísel.
- U barevných obrazů se využívá principu, že každý barevný odstín lze získat smícháním tří základních barev v určitém poměru:
 - Červené (**R**ed), zelené (**G**reen) a modré (**B**lue).
- Tento model skládání barev se nazývá **RGB model**.
- Pro něj platí, že smícháním všech tří barev s plnou intenzitou získáme bílou barvu. Intenzita jednotlivých barevných složek se vyjadřuje pomocí čísla.
- To je obvykle osmibitové, tzn. v rozsahu 0-255. Znamená to, že každá ze složek může nabývat dvě na osmou, tedy 256 různých hodnot jasu.

Ukázka barev RGB



Formáty digitalizovaných materiálů

- Obraz lze uložit v mnoha různých formátech.
- Ty mají obvykle za úkol navíc nějakým způsobem redukovat množství dat potřebných k popisu obrazu.
- V tomto případě jde o tzv. komprimované obrazové soubory.
- **Podle typu komprese je lze rozdělit:**
 - na **bezztrátové** (po komprimaci získáme obraz naprosto shodný s původním)
 - na **ztrátové** (po komprimaci získáme obraz velmi podobný původnímu, přičemž ztráta je obvykle volena taková, aby nebyla postřehnutelná)
- **Typy formátů podle názvu:**
 - Nekomprimované nebo s bezztrátovou kompresí: **BMP, TIFF**
 - Se ztrátovou kompresí: **JPG**

Digitalizace

- Papírový dokument je stále nejrozšířenější forma pro udržení archivních informací.
- Papírový dokument klade velké nároky na prostor, neboť počet dokumentů se neustále zvětšuje.
- Se vzrůstajícím množstvím se objevuje ještě další nevýhoda tohoto zpracování.
- **Vyhledání** více informací podle nějakého kritéria je téměř nemožné a nebo hodně zdlouhavé.
- V dnešní, počítači zahlcené době, se přímo nabízí počítačové zpracování a uchování všech dokumentů. Skenování umožňuje převést papírový dokument do digitální podoby.

Papírový dokument

- Nejrozšířenější forma pro udržení archivních informací.
- Klade velké nároky na prostor, neboť počet dokumentů se neustále zvětšuje.
- Vyhledání více informací podle nějakého kritéria je téměř nemožné anebo hodně zdlouhavé.
- Řešení:
- Převést papírový dokument do digitální podoby = digitalizovat
- Využít skenování.

Výhody digitalizace

- Zrychlení administrativní i správní činnosti (rychlý přístup občanů k informacím).
- Snížení rizika zničení, porušení, ztráty či ztráty dokumentů.
- Rychlé prohledávání digitálního archivu.
- Možnost elektronické distribuce a prezentace (intranet, internet).
- Snížení prostorových nároků a možnost přesunu archivovaných originálů na méně exponované lokality.
- Kompletní prověření a "inventura" digitalizovaného dokumentového fondu
- Nejdůležitější však je vytvoření základního předpokladu pro zlepšení kvality přístupu k informacím pro občany.

Postup při digitalizaci

- Skládá z několika pracovních postupů, které na sebe musí plynule navazovat.
- Základní pracovní postupy jsou:
 - ▣ Příprava ke zpracování
 - ▣ Skenování
 - ▣ Indexace
 - ▣ Verifikace
 - ▣ Zařazení do digitálního archivu
 - ▣ Kompletace po zpracování

Metoda digitalizace

- Při digitalizaci je nutné zamyšlení, jaký bude mít digitalizace postup a potom i celý výsledek procesu.
- Papírové spisy neustále přibývají a jistě dojde k situaci, kdy se nový dokument bude muset zařadit do již zdigitalizované části kanceláře nebo archivu.
- Aby se předešlo nekonzistentním informacím, je nezbytné zvolit správný postup.
- Nabízí se řešení paralelního přístupu.
- Skenování lze postavit na dvou liniích zpracování.
 - Skenování postupně podle druhu dokumentů, tak jak jsou roztrženy.
 - Skenování nových dokumentů.

Postup - Příprava ke zpracování

- Před vlastním zpracováním je třeba připravit dokumenty tak, aby digitalizace jako celek probíhala co nejefektivněji, bezchybně a bez problémů.
- Příprava papíru pro skenování
 - ▣ Odstraňování sponek a svorek.
 - ▣ Rovnají dokumentů.
 - ▣ Zalepení dokumentů v případě roztrhnutého papíru.
- Příprava pro další zpracování
 - ▣ Vkládání speciálních papírů pro oddělení rozešitých dokumentů
 - ▣ Tvorba čárových kódů, které pak usnadní indexaci.
 - ▣ Lepení nálepek s čárovými kódy.

Skenování

- Samotné skenování nečiní obvykle problémy.
- Při skenování vlastními silami je předpokladem úspěchu nejen dobrá příprava, ale také zaškolení pracovníků, kteří skener (skenery) obsluhují.
- Velkou pozornost je nutné věnovat nejen výběru skeneru, ale také výběru software pro skenování a manipulaci s obrazem.
- Zde jen může vzniknout riziko poškození původního materiálu v důsledku špatného výběru skenovacího zařízení.

Indexace a metadata

- Indexace je nedílnou součástí zpracování.
- Samotný obraz dokumentu, byť sebelépe naskenovaný, není možné rychle v digitálním archivu vyhledat, pokud s ním není spojena dostatečně vypovídající hodnota indexu.
- Jako index může být použit jakýkoli individuální charakteristický znak dokumentu - podle druhu agendy může být indexem například *rodné číslo, příjmení, číslo smlouvy, číslo popisné, číslo katastru*, atd.
- Každý dokument má potom ještě další charakteristiky v podobě metadat.

Verifikace

- V některých případech je třeba ještě provést kontrolu kvality digitálního obrazu, aby se mohl digitální formát obraz pořídit popřípadě ještě jednou.

Zařazení do digitálního archivu

- Každý digitální archiv má svoji pevně danou strukturu, aby bylo možné dokumenty vyhledávat.
- Někdy přesně kopíruje uspořádání papírové ho archivu, jindy je šikovnější vytvořit jiné.
- S uspořádáním dokumentů je třeba většinou soubory přenést na jiné místo, nejčastěji na diskové pole.
- Toto kopírování může mít za následek nemalé zatížení sítě.
- Velké výkresy s vysokým rozlišením v dobré kvalitě mohou zabrat poměrně dost místa, řádově stovky MB

Kompletace po zpracování

- Po zpracování je nutné uvést zpracované předlohy do původního stavu, popřípadě do stavu požadovaného.
- Zpravidla je nutné znovu sešít svorkou dokumenty, které byly během přípravy rozešity.
- Některé dokumenty je žádoucí uložit do speciálních obalů, často je žádoucí použít jiný typ archivačních krabice, než byly původní, může být použit jiný způsob značení, atd.

Návrh struktury digitálního archivu

- Každý dokument v archivu nese informace, které jsou důležité pro jeho zařazení a pozdější vyhledání.
- Z těchto důvodů je třeba vytvořit správné uspořádání a označení dokumentů.
- Tyto informace je třeba zjistit od pracovníků v archivu nebo osob, které budou digitální archiv využívat.
- Uspořádání je obvykle provedeno do několika úrovní složek, kde jsou potom uloženy samotné dokumenty popsané **metadat**.
- Dokumentů může být několik typů podle informace, kterou obsahují.
- Metadata jsou uspořádána do evidenčních karet, které obsahují všechny popisné informace o dokumentu.

Příklad ze stavebního archivu

- ***Základní strukturální elementy stavebního archivu:***
 - Katastrální území - je druhou nejvyšší úrovní struktury úložiště. Jedná se o skupinu jednotlivých katastrálních území městské části Prahy 10.
 - Stavba - stavba je další úrovní struktury úložiště. Může se jednat o budovu, liniovou stavbu, případně jiný typ stavby.
 - Spis - spis je nejbližší nižší úrovní organizační struktury úložiště. Spis charakterizuje jednotlivá řízení vztahující se k příslušné stavbě.
- ***Základní typy dokumentů stavebního archivu:***
 - Dokument - nejdůležitější typ dokumentu, který s sebou nese, o jaké rozhodnutí se jedná. Je určen jednacím číslem.
 - Výkres - grafický doprovodný dokument obsahující především výkresovou dokumentaci
 - Doklady - doklad je listina, která je potřeba k vydání rozhodnutí nebo mu předchází, případně se k němu vztahuje jiným způsobem

Metadata elementu stavba

- **karta OST stavba**
 - město
 - katastr
 - číslo popisné
 - číslo parcelní
 - číslo evidenční
 - číslo orientační
 - číslo nouzové
 - ulice 1
 - ulice 2

Struktura dat příklad

- 1. úroveň: město - Praha 10
- 2. úroveň: katastrální území
- 3. úroveň: stavba
- 4. úroveň: spis
- 5. úroveň: jednotlivé dokumenty

Skenovací zařízení - Parametry scannerů

- Skenování - Při skenování předlohy scanner pomalu osvětluje a snímá (podobně jako třeba kopírka) předlohu. Snímají se po řadách jednotlivé body, ze kterých se výsledný obraz bude skládat (protože jsou tyto body velmi malé, lidské oko je nedokáže rozlišit a výsledek vypadá celistvě).
- **Základní parametry scannerů:**
- Barevná hloubka
- Rozlišení obrazu
- Maximální velikost snímaného obrazu

Barevná hloubka

- Udává množství odstínů barev, které je schopen scanner nasnímat.
- Dnes obvyklou barevnou hloubkou je 24 bitů, což znamená možnost záznamu v 16 777 216 odstínech.
- U profesionálních přístrojů dosahuje barevná hloubka až 48 bitů (281 474 976 710 655 odstínů).

Rozlišení obrazu

- Udává se obvykle v DPI (počet tiskových bodů na palec) a znamená jemnost snímacího rastru a potažmo s tím spojenou datovou velikost výsledného obrazu.
- S větším rozlišením se tato velikost zvyšuje, ale s tím roste i kvalita digitálního dokumentu

Maximální velikost snímaného obrazu

- Každý skener má jako jeden ze svých parametrů maximální velikost snímané předlohy.
- Většina běžných skenerů umožňuje skenovat formáty A4.
- Velkoformátové skenery umožňují skenovat až formáty A0.
- Umí-li skener větší formát, bude už jistě daleko dražší.

Výběr zařízení

- Automatický podavač
- Ploché lože
- Velkoformátové skenování
- Skenování vázaných předloh

Automatický podavač

- Rychlost skenování je vždy nejvyšší tehdy, jestliže se využívá automatický podavač předloh.
- Podle kvality předlohy se v tomto případě vkládají předlohy do podavače nebo jeho štěrbinu po jedné (při nízké kvalitě předlohy) nebo se využije zásobník (podle druhu skeneru na cca 100 až 500 papírů).
- Současné skenery jsou vyráběny a dodávány vždy s automatickým podavačem předloh ("ADF" - Automatic Dokument Feeder).

Ploché lože

- Pomalé je použití tzv. plochého lože, kdy je nutné, podobně jako u běžné kopírky, položit předlohu na skleněnou podložku.
- V tomto případě jde nejen o pomalejší manipulaci v porovnání s použitím podavače, výrazně pomalejší je i samotné skenování.
- Skenování je v tomto případě daleko šetrnější k poničením dokumentům.
- Ne všechny skenery jsou však možností skenovat dokumenty z plochého lože ("Flat Bed") vybaveny.

Velkoformátové skenování

- Při skenování velkých formátů (větších než A3) je nutné použít speciální druh skenerů.
- Rychlost skenování může být relativně vysoká a může odpovídat rychlosti skenování z plochého lože.
- Vzhledem k tomu, že délka skenované předlohy je však vyšší, prodlužuje se odpovídajícím způsobem také doba skenování.
- Při skenování starých a málo soudržných předloh velkých formátů je nutné před skenováním nejprve vložit předlohu do speciální průhledné folie.
- Tato manipulace má samozřejmě také vliv na dobu skenování.

Skenování vázaných předloh

- Pro skenování vázaných předloh jsou určeny speciální, **tzv. knižní skenery**.
- Mezi jejich typické vlastnosti patří, že předloha je skenována bezkontaktním způsobem "zhora", skener je často vybaven zařízením na obracení stránek a speciální technologie a korekcí optického zkreslení obrazu u hřbetu předlohy.
- Skenování jednotlivých stran na knižním skeneru je výrazně rychlejší než při skenování z plochého lože.

Skenery

- **Velkoformátové skenery**
 - HP - DesignJet Scanner 4500 A0+, Canon - Colortrac SmartLF Cx40e, Contex - HAWK-EYE Cx 36,
- **Skenery pro nejvyšší zátěž – až 200 stránek za minutu**
 - Kodak i1860, Kodak i1840, Kodak i800
- **Skenery pro vysokou zátěž – až 130 stránek za minutu**
 - Kodak i780, Kodak i600
- **Skenery pro střední zátěž**
 - Kodak i1400, Kodak i1300, Kodak i280, Kodak i250 & 260, Fujitsu fi-5750C, Fujitsu fi-5650C
- **Skenery pro nízkou zátěž**
 - Kodak i1200, Kodak i1120, Kodak i150 & i160, Kodak Scan Station 100, Kodak i55 & i65, Kodak i30 & i40, Fujitsu fi-4340C

Canon Colortrac SmartLF Cx40e

- používá snímače CIS se skutečným optickým rozlišením 600 dpi.



HP Designjet 4500

- rychlost barevného skenování je až 76 mm za sekundu
- rychlost černobílého skenování až 254 mm za sekundu



Kodak i780

- až do rozlišení 300 dpi rychlost 130 stránek/min



DMS systémy pro ukládání dat

- **Document management systems** lze dnes definovat jako počítačové systémy (programy) používané k ukládání, uchovávání, archivaci a sdílení elektronických dokumentů.
- Může jít o texty nebo obrazové dokumenty, které vznikly již jako elektronické nebo byly do této podoby převedeny.

Základní funkce systémů DMS

- Vkládání dokumentů
- Zabezpečení
- Nastavení práv přístupu
- Organizování dokumentů
- Verzování dokumentů
- Workflow
- Archiv
- Zálohování

Vkládání dokumentů

- Všechny systémy DMS disponují možností nahrávání dokumentů, většinou ve formě jednoduchého uploadovacího formuláře.
- Pokročilejší systémy jsou vybaveny OCR softwarem, který umožňuje rozpoznání obsahu a jeho indexaci pro fulltextové vyhledávání.
- Při vkládání jakéhokoliv dokumentu je možné doplnit metadata (název, autor, popisek, klíčová slova apod.), systém pak doplňuje další údaje o dokumentu jako jeho velikost, formát, datum vložení a uživatele, který dokument vložil.

Zabezpečení

- Patří k nejdůležitějším součástem DMS systémů.
- Požadavky na zabezpečení by měly hrát hlavní roli při výběru konkrétního řešení.
- Je těžké stanovit, jaká má být nejnižší úroveň zabezpečení.
- Vždy závisí na důležitosti dat.
- Se zabezpečením úzce souvisí také problematika přístupových práv.

Nastavení práv přístupu

- Tato funkce je závislá na složitosti systému a samozřejmě potřebách organizace.
- V zásadě by mělo fungovat alespoň základní přidělování práv k souboru, tzn. kdo jej má právo číst, editovat nebo mazat.
- Jednotlivá práva mohou být nastavena určité skupině uživatelů nebo v pokročilejších systémech přímo konkrétním osobám

Organizování dokumentů

- Cílem jakéhokoliv informačního systému je nalézt vždy to, co potřebujeme.
- Základem každého DMS by měl být systém složek, které slouží k organizování dokumentů.
- Kvalitnější DMS k tomu přidávají také vyhledávání a to zejména fulltextové.
- K jeho provozování je zapotřebí kvalitní indexace a metadatový popis na vstupu.

Verzování dokumentů

- Tato funkce umožňuje v systému ukládat všechny změny v dokumentu.
- Kdykoliv v budoucnosti je pak možné vrátit se k některé z předchozích verzí.
- Dojde-li k nechtěnému smazání části nebo celého dokumentu, případně k nějaké nežádoucí změně, lze jednoduše nahrát některou ze starších verzí.
- Navíc u každé změny se eviduje její datum a také uživatel, který změnu provedl.

Workflow

- Tato funkce patří již spíše k těm pokročilejším a bývá přizpůsobena na míru konkrétní firmě.
- V rámci každé instituce procházejí dokumenty určitým informačním tokem, tzn. kolují mezi zaměstnanci nebo odděleními.
- Je-li tato funkce v systému DMS zastoupena, můžeme dokumentu nastavit jak, kdy a komu se má zobrazit.
- Lze tedy přesně definovat, kdo může dokument upravovat, kdo a na jaké úrovni má dokument schválit, případně kde a kdy se má zveřejnit.

Archiv

- V každé firmě nalezneme různě rozsáhlé archivy.
- Trendem dneška je jejich digitalizace a ukládání v elektronických repozitářích.
- A právě ty jsou další součástí systémů DMS.
- Mohou uchovávat nejen zdigitalizované dokumenty, ale samozřejmě do nich můžeme přesunovat vyřízenou elektronickou agendu.
- Výhodou je výrazná úspora místa, jednodušší vyhledávání dokumentů a samozřejmě také jednodušší manipulace s nimi.

Zálohování

- Každý elektronický systém musí mít vyřešenu otázku zálohování dat a případně jejich migraci do jiných perspektivnějších formátů pro udržení kontinuity přístupu k nim.
- Nejinak je tomu i v systémech DMS.
- Zálohování by mělo být pravidelné a administrátor by měl být vždy schopen rychle nahrát data ze starší zálohy.
- I pro zálohy platí bezpečnostní opatření, tzn. měly by se k nim dostat vždy jen kompetentní osoby.

ELO Professional

- Základní vlastnosti:
 - ▣ Určeno pro platformu Microsoft Windows.
 - ▣ Využívá databáze SQL Microsoft (včetně MS SQL 200 Express) nebo Oracle.
 - ▣ Určeno pro 3-50 současně pracujících uživatelů.
 - ▣ Licence pro pasivní uživatel zdarma.
 - ▣ Fulltextové vyhledávání.
 - ▣ Verzování všech typů dokumentů.
 - ▣ Workflow v základní konfiguraci.
 - ▣ Napojení na aplikace MS OFFICE (Word, Excel, PowerPoint, Outlook).
 - ▣ Napojení na Internet Explorer.
 - ▣ Optimalizace skenovaných dokumentů.

EMC Documentum

- Documentum 6 obsahuje komponenty:
 - **Enterprise Document Management (EDM)** – Správa dokumentu, jak papírových originálů, tak elektronických v agendách jako jsou došlé faktury, směrnice, smlouvy, nabídky apod.
 - **Web Content Management (WCM)** – Tvorba, správa a publikace obsahu webu známá také jako redakční a publikační systémy
 - **Digital Asset Management (DAM)** – Správa a zpracování multimediálních souborů jako jsou obrázky, zvuk, video
 - **Business Process Management (BPM)** – Návrh, modelování a řízení business procesů nejen s dokumenty, ale i nad rámec úložiště včetně integrace okolních systémů, aplikací a obsahu
 - **Collaboration Content Management (CCM)** – Řízení a podpora operativní spolupráce projektových týmů
 - **Records Management (RM)** – Archivace závazných dokladů,
 - **Compliance Content Management (CCM)** – Tvorba a řízení závazné řízení dokumentace.

Jiné aplikace - OCR

- **OCR** neboli optické rozpoznávání znaků (Optical Character Recognition) je metoda, která pomocí scanneru umožňuje digitalizaci tištěných textů, s nimiž pak lze pracovat jako s normálním počítačovým textem.
- Počítačový program převádí obraz buď automaticky nebo se musí naučit rozpoznávat znaky.
- Převedený text je téměř vždy v závislosti na kvalitě předlohy třeba podrobit důkladné **korektuře**, protože OCR program nerozezná všechna písmena správně.
- Je použitelné pro všechny tištěné výstupy z laserových, inkoustových, termosublimačních a jehličkových tiskáren a samozřejmě pro předlohy vytištěné knihtiskem.
- **U nevhodných předloh např. slabě vytištěných jehličkových tiskáren nebo dohromady slitých písmen se z časového hlediska vyplatí spíše přepis textu.**

Digitalizace ve státní správě

- Digitalizace je využívána hlavně v rozsáhlých archivech. Jde především pro **stavební archiv**, které obsahují různorodé materiály.
- Dost často se digitalizuje i **živnostenský archiv**, protože je to relativně nový archiv, který obsahuje dokumenty vytvořené až po roce 1989.

Projekty digitalizací v městských částech Prahy

□ Praha 6

- Městská část Praha 6 je s digitalizací nejdále, začali v roce 2007, kdy se předpokládalo, že se do digitální podoby převede 6,5 miliónu předloh a dokumentů. Předpokládalo se, že proces digitalizace bude celkem stát zhruba 15 mil. korun.

□ Praha 2

- Městská část Praha 2 má za sebou pilotní projekt.

□ Praha 10

- Městská část Praha 10 má pilotní projekt a v současné době je vypsáno výběrové řízení na digitalizaci stavebního archívu.

□ Praha 8

- Městská část Praha 8 se snaží o digitalizaci vlastními silami, ale ukládání metadat do tabulky excelu se brzy ukázalo jako nevhodné. V současné době se pracuje na výběrovém řízení pro digitalizaci stavebního archívu.

Zdroje informací

- JEDLIČKOVÁ, Petra. Multimédia: Úvod do digitalizace [online]. Praha: Univerzita Karlova,
- Filozofická fakulta, Ústav informačních studií a knihovnictví, 2002. [cit. 2004-04-30].
- Studijní materiál k semináři Multimédia. Dostupné z WWW:
- http://http://web.ff.cuni.cz/~jedlickp/kurzy/multimedia1_2/multimedia5_2002/index.htm
- PECINOVSKÝ, Josef. Skenery a skenování. 2. aktualiz. vyd. Praha: Grada, 2003. 120 s.
- Snadno a rychle. ISBN 80-247-0544-3.
- Internetové portály městských částí www.praha2.cz, www.praha6.cz

Konec

- Děkuji Vám za pozornost.
- Kristýna Knapová