

# SVE VIŠE PODATAKA – SVE MANJE PROSTORA

Miroslav Milovanović, Mfc&L d.o.o.

## *Sadržaj*

*Članak prikazuje probleme kod procesa digitalizacije, :*

- *Informacijska tehnologija*
  - *Važnost se daje trenutačnoj obliki digitalizacije i procesima koji se upotrebljavaju*
- *Kvaliteta arhivske materije*
  - *Specificiranje materije koja se digitalizira u procesu obrade i naručnikove zahtjeve prilikom poslovanja..*
- *Problemi kod skladištenja*
  - *Problemi, koji se pojavljuju kod skladištenja podataka u postojećim informacijskim infrastrukturama i problemi, koji još dolaze.*
- *Prostorske rešenja*
  - *Nekoliko rešenja sa kojim možemo racionalizirati poslovanje unutar sofisticiranog poslovnog sistema.*

## *Abstract*

*The article is trying to reveal certain problems within the digitalizing process. We could separate the content to following parts:*

- *Informational technology*
  - *Emphasis is on already defined ways of digitalizing documents and processes that are used.*
- *Quality of material*
  - *Defining the state of the material is one of the crucial parts of the digitalization, and presenting the customers terms within the process.*
- *Problems with saving*
  - *Problems that are present in the current informational structure and the problems that are jet to come.*
- *Solutions for efficient saving*
  - *Some of the solutions that are used to prevent unnecessary space usage and some solutions for optimizing business process.*

Elektronsko arhiviranje je područje, koje sa uvek novim adaptacijama i dopunjivanjem nudi preglednije i racionalnije poslovanje. Svedoci smo ubrzanju digitalizaciji arhivskog gradiva tokom poslednjih godina. Papirnati dokumenti nisu jedino gradivo koje se ubrzano digitalizira, sve aktivnije se digitaliziraju i drugi analogni oblici (audio, video zapisi,...).

U procesu arhiviranja se kod organizacije uvek pojavljuju novi izazovi, kako klasičan arhiv efikasno pretvoriti u digitalni oblik i kako skladištiti podatke da su ubuduće dostupni naručnicima. Od svih područja digitalnog arhiviranja među ostalima istupa i prostorska politika kojom se susreću sve više organizacija koje vrše digitalizaciju arhiva .

Sve veći broj dokumenata, a i zahtevniji uslovi zahvate podataka diktiraju dimenziju digitalnog oblika, a to posledično zahteva sve više prostora za arhiviranje. Rešavanje problema se u većini primera oslanja na odnos između veličine i kvalitete konačnog rezultata.

Razvoj tehnologije za arhiviranje

Tehnologija koja se upotrebljava unutar digitalizacije arhiva je uveliko različita od materiala, koji se arhivira. Zbog prilagođavanja i optimizacije poslovanja preduzeća u većini razvijaju vlastita programska rešenja.

U preduzeću MFC&L smo uoči sve većih zahteva naručnika uveli vlastito programsko rešenje ADS (Arhivsko dokumentacijski sistem), koje omogućuje brzo prilagođavanje zahtevama naručnika. Komercijalne verzije arhivskih programa nam nisu omogućavale brzo prilagođavanje niti preciznosti kod potencialnog posla, dok međutim preko vlastitog programskog rešenja možemo reagirati brzo i usmerivati rešenja u korist naručnika.

Kod tehnologije koja se upotrebljava pri procesu digitalizacije, bitno je izpostaviti sledeće oblike digitalizacije:

## **OCR (Optical character recognition).**

Kod OCR tehnologije se dokument iz klasično grafičkog zahvata digitalno obrađuje tako, da sistem obavi prepoznavanje znakova i u konačnom obliku dokument posreduje u tekstovnom formatu.

## **OMR (Optical Mark Recognition).**

Kod OMR tehnologije sistem kod npr. skeniranja čita vrednost ili simbole ili oznake na za to unapred pripremljenom standardiziranom obliku dokumenata (prijavnice na fakultetama, ankete, bankovna dokumentacija – Slovenija). U tom primeru program čita samo delove dokumenata (okviri). Kod OMR tehnologije prepoznavanje ima samo dve izlazne vrednosti a to su ili je izabrano ili nije. Unutar konfiguracije radnih modela sistem se može prilagoditi na način, koji omogućava dodatnu analizu podataka (statistika) i ne samo indeksnu vrednost. Uglavnom se OMR upotrebljava za konsistentnu obradu podataka na osnovu standarda ili oblika, koje su usvojene unutar specifične organizacije. Kao primer OMR tehnologije je bar koda najrazširenija oblika prepoznavanja identifikacijskih podataka koja se danas upotrebljava.

## **ICR (Intelligent Character Recognition).**

ICR tehnologija je tehnologija, koja je još u razvoju ali koja nudi uz standardno odčitavanje simbola i odčitavanje rukom pisanih tekstova. U ovom trenutku je rešenje još uvek ograničeno na odčitavanje velikih slova. Taj projekat nudi rešenje za sve, kojima je onemogućena normalna upotreba savremenih tehnoloških sredstava.

Kod gore opisanih metoda digitalizacije u prostorskom arhiviranju nema većih problema jer je razmerje veličine ulazne i izlazne oblike, to jest veličine slike (input) i tekst datoteke (output) prilično veliko tako da se posledično time olakša odluka za konačan oblik.

## **Slikovna obrada**

Kod slikovne obrade je u većini primera mišljena digitalizacija klasičnog arhiva, kao transformacija iz digitalno-tekstovnog oblika ili zahvate podataka skeniranjem.

Razlika je u tome, da je konverzija kod digitalnih podataka racionalnija jer kod zahvata ulaznih podataka u obliku skeniranja dolazi do mnoštva prostorsko opterećujućih faktora (senke, nepotrebni podaci, rezolucija)

Slikovna obrada jedan je od najrazširenijih i najprihvatljivijih formata digitalizacije klasičnog arhiva, ali je zbog sve većih zahteva po boljoj kvaliteti i prostorski opterećujuća.

## VR (Voice Recognition)

Protoklih nekoliko godina počela se je razvijati tehnologija za prepoznavanje govora i time automatizam konverzije iz audio formata u tekstovni oblik. Ponudnika na tom području nema puno (Dragon Naturally Speaking,...), problem je i u nedopunjavanju baze stranih jezika. Ovakva rešenja više spadaju u delokrug optimizacije poslovanja ali ne treba zanemariti učinak konverzije iz audio u tekstovni oblik, gde je vidna razlika u konzumaciji prostora.

Kvaliteta arhivskog gradiva i kvaliteta obrade (stanje originala i rezultati obrade)

Proces obrade dokumentarnog materiala počinje sa pregledom dokumentacije kod kojeg je sve ovisno o brzini, veličini i preciznosti zahvate podataka. Sam pregled se razlikuje od tehnologije, koja je upotrebljena za obradu dokumenata. Dok neki organizacijski sređuju materiju tako, da je pred obradom sređena tako da ima i u digitalnom obliku linearni i pregledni put, drugi obrađuju dokumente bez potrebnih ključeva za sređivanje koje dodeljuju naknadno pomoću indeksnih atributa.

Stanje arhivske materije je od iznimne važnosti za kvalitetnu obradu i kvalitetnu digitalnu obliku. Ocena izvedljivosti zahvata podataka u digitalan oblik uslov je za prilagođavanje daljeg procesa na svim područjima kao što su hardver, softver i organizacija rada. Dokumentima koji nisu u stanju za normalnu obradu potrebno je dati oblik koji neće osporavati normalan radni proces a koji pogotovu omogućuje preglednost obrade sa ostalim podacima i njihovu skladnost.

Kvaliteta obrade je rezultat digitalizacije dokumenata, a njihov izlazni oblik se razlikuje od uslova postavljenih pred samom obradom.

Kvaliteta obrade definisana je u većini slučajeva sa dva uslova:

- Tehničke pretpostavke zahvate podataka (unutar tih pretpostavki naručnik određuje kakvu obliku mora imati dokument kad prolazi kroz proces obrade, npr. skeniranje dokumenata u specifičnoj rezoluciji, format datoteke, kompresija,...).
- Tehnologija, koja se upotrebljava kod obrade dokumentarne materije. Tehnologija je onaj element koji omogućava konkurenčnu prednost, jer već sada pratimo razvoj produkata, koji u većini podpostupka digitalnog arhiviranja nude celovito rešenje.

Gore opisani uslovi nisu jedini koje definišu ili utiču na konačni oblik digitalne dokumentarne materije jer je kod tako kompleksnog procesa već u početku ovisno i u kakvom stanju je materija kada je predana obradi. Za većinu dokumentacije bi mogli reći da je ulazni oblik jednak izlaznom ili jednostavnije ako je dokument u slabom stanju onda se ne može omogućiti delova kojih nema. U nekim primerima je dozvoljena manipulacija podataka zbog osiguravanja kvalitete (ta manipulacija se nanosi uglavnom na dokumentaciju kod koje je u interesu naručnika da se uklone greške, npr. šumovi kod audio zapisa).

Problemi kod skladištenja podataka (prostorski i ostali)

Problemi kod skladištenja nisu vezani samo na sam prostor nego i na način kako omogućiti uvid u dokumente a i na brzinu pregledivanja samog dokumenta.

Živimo u doba globalnih mreža, gde se sa razvojem tehnologije nudi i dostup do materije koju je potrebno videti i sa odaljenih lokacija.

Cilj je omogućiti uvid u dokumentaciju ili arhiv na način, koji omogućuje naručiocu iz odaljene lokacije brz dostup i lako pronalaženje sadržaja arhiva. Problemi koji se ukazuju kod skladištenja digitalnih dokumenata su uvek vezani na prostor gde se sladište.

Ako uzmemo na primer, da arhiviramo skenirane dokumente u A0 formatu, a prihvaćen je zahtev naručioca da se ne upotrebljava kompresija kod skladištenja, onda može jedna takva datoteka da dostigne veličinu od 500MB i više (rezolucija 400dpi). Ako je govora o malom broju dokumenata onda se za takav arhiv upotrebe uobičajeni nosioci podataka (cd, dvd, diskovna polja...). Problem

nastaje kada treba na zahtev naručnika ohraniti originalne datoteke, a usto omogućiti dostup i brzo pregledivanje materije (A0 format). Kada imamo u mislima nekoliko hiljada takvih datoteka onda govorimo i o više TB prostora za skladištenje tih dokumenata.

Problem koji se ukazuje kod gore navedenog procesa možda i nije toliko problematičan sa vidika osiguravanja prostora, ali se zaoštruje kad moramo do tih datoteka omogućiti dostup i uvid preko elektronskog arhiva a pogotovu preko mreža. Samo arhiviranje može biti problematično već zbog uzdržavanja u primeru da posedujemo sisteme, koji nam ne dozvoljavaju menjanje ili brisanje podataka. Ako arhiviramo podatke za koje nemamo zakonskog ograničenja trajanja a imamo zahtev naručioca, da se mora upotrebiti tehnologija koja onemogućava brisanje ili menjanje podataka, onda je prostorsko gledano toliko veći rizik za očuvanje takvih podataka jer nemamo garancije da će biti naručilac poslovno aktivan toliko vremena za koliko su arhivirani podaci.

Rešenja protiv zauzimanja prostora

Rešenja su različita glede na nameru i predvidljive upotrebe digitaliziranih dokumenata. Kao što je već bilo pomenuto, potrebno je klasificirati da li će biti dokumenti arhivirani samo u elektronskom obliku ili je proces digitalizacije servisna usluga, kojom omogućavamo uvid u te digitalne dokumente.

Poznajemo više rešenja kojima rešavamo takve probleme.

- Izdvajanje nepotrebne, neupotrebljive, nevažne, podvojene dokumentacije ili dokumentacije koje je zakonom dozvoljeno uništiti iz klasičnog i digitalnog arhiva.

Tokom digitalizacije je prvo rešenje provera materiala, koje je određeno za pretvorbu u elektronski format. S tim izbegnemo nepotrebno podvajanje neupotrebljivih i zakonito zastaranih dokumenata.

- Sistemska rešenja, koja nam omogućuju pojedinačnost zapisa (rešenje koje je možda važnije za preglednost digitalnih dokumenata, jer se time smanjuje dodavanje identifikacijskih elemenata).
- Digitalni formati sa velikim faktorom kompresije datoteka, koji ohranjuju visok kvalitet podataka (lossless compression), npr. DjVu format.

Kompresija podataka je važan deo digitalizacije, jer se tu kao konačni deo procesa odlučuje veličina datoteke s obzirom na kvalitetu kompresije. Od velike važnosti je i oblikovanje što manje datoteke zbog dostupa preko mreža (internet). Ako bi hteli preneti ili pogledati datoteku npr. tiff – 500MB ili jpeg – 40MB to u mnogočemu predstavlja problem, jer ne samo, da takvi prenosi mogu da potraju nego je i upitljiva izvedba takvih operacija na nekim infrastrukturama. Jedno od rešenja koje je bilo razvito s namerom prenosa preko mreža je DjVu kompresija ili DjVu format. Format koji omogućuje izuzetno jaku kompresiju bez ili sa minimalnim gubitkom podataka postaje prednostni oblik kod prikaza grafičnih datoteka preko interneta. Pri tome je uvek treba paziti na specifikacije formata koje je naručio naručnik.

- Odabir formata u koje se pretvara dokumentarna materija.

Odabir formata je isto tako važan korak digitalizacije. Ograničenja kod biranja formata su najčešće pravna ili vezana na standarde jer ti određuju način i oblik kojim neka organizacija posluje (npr. ako želimo, da je digitalni dokument usklađen sa poslovanjem nekog drugog sistema onda moramo zagotoviti prepoznatljivost takvih dokumenata na drugom sistemu).

Navedena rešenja su samo neka od mnogih. Svaka organizacija usvaja rešenja u okviru vlastite informacijske infrastrukture i vizije poslovanja.

S obzirom na zahteve po kvalitetnoj obradi klasičnih dokumenata ili dokumenata izrađenih u digitalnom obliku i s obzirom na porast poslovne dokumentacije, raste i potreba za sigurnim prostorom skladištenja podataka. Informacijska tehnologija nam sa svojim razvojem nudi rešenja koja se globalno još dalje razvijaju.

