

# Aktuelni problemi digitalizacije u državnoj administraciji

Nikola Popović, Julijana Mirčevski

**Apstrakt** --Tekuće stanje procesa digitalizacije u Srbiji počinje da ispoljava prve, jasne znakove protivrečnih pojava. Uvodi se značajan broj sistema za upravljanje dokumentima, posebno u sistemu državne administracije. Uz sasvim očite pogodnosti, ispoljavaju se problemi kompatibilnosti novih sistema sa postojećim sistemom rada sa papirima. Harmonizacija dostupnosti dokumenata i načina rada je otežana. Jedan deo dokumenata i pored digitalne verzije mora i dalje da ostane na papiru. Troškovi digitalizacije su visoki, a navedeni problemi izazivaju značajna kašnjenja implementacije. U radu su naznačena moguća tehnološka rešenja koja omogućavaju održivost procesa digitalizacije kroz niže troškove.

**Ključne reči** -- digitalizacija; održivost; metodologija digitalizacije; državna uprava.

## I. UVOD

Analiza metapodataka koji su u propisima predviđeni za opis digitalnih dokumenata ukazala je da je strategija digitalizacije zasnovana na klasičnim principima rada sa papirom. Regulativu u većini čine propisi koji se tiču zaštite digitalnih dokumenata. Usvojen je klasični princip grupisanja dokumenata hijerarhijskom klasifikacionom šemom. Nismo uočili da se iko bavio sadržajem dokumenata, semantikom i odgovarajućim aspektima jezika.

Težište digitalizacije je usmereno na zadovoljavanje potreba građana i preduzetnika. Od državnih organa se primarno zahteva i očekuje znatno veća efikasnost i uštede. Masovno se pominje primena "najboljih praksi" (eng. best practices), ali bez navodjenja čije prakse i konteksta te prakse. Proces izrade studija stanja (eng. feasibility study) se sprovode/zamenjuju prostim anketiranjem. Izučavanje kompleksnih, heterogenih procesa u državnoj administraciji ne može da se izvodi anketiranjem. Unifikacija radnih procesa je nekada opravdana, ali u nekim nedovoljno analiziranim situacijama može da dovode do gubitka bitnih podataka i nemogućnosti generisanja novih informacija. Ciljevi digitalizacije, kako su postavljeni, ne uzimaju dovoljno u obzir potrebe organizacije rada u državnim organima. Regulativa je orijentisana na što bržu digitalizaciju u domenu finansija i bankarstva. U tom domenu najveći procenat dokumenata čine strukturirani dokumenti. Državna administracija u najvećem procentu radi sa nestruktuiranim dokumentima. Proces rada sa struktuiranim i nestruktuiranim dokumentima se suštinski

razlikuju. Radni procesi u državnim organima često zahtevaju organizovanje kompleksa podataka i metapodataka, koji omogućava složenu konsolidaciju informacionih resursa, u cilju donošenja racionalnih dokumentovanih odluka. U takvim slučajevima često je nužno formiranje ad-hoc klastera dokumenata radi sagledavanja sadržaja i obima problema. Jednostavan primer je ažuriranje podsetnika za posetu ministra stranoj zemlji. Osim novih podataka potrebno je proveriti i revidirati arhivske podatke (istorijat) radi njihovog sagledavanja (prikazivanja ili ne) u kontekstu tekućih zbivanja. Ovi postupci uslovljavaju pristup mnogo većem broju dokumenata, u vremenskom i tematskom domenu. U državnoj administraciji masovno se uvodi "Document management system" - DMS. DMS po definiciji ima sam dokument u centru pažnje (eng. document centric approach), što znači da ne odgovara uvek značajnim procesima informisanja u državnoj administraciji.

## II. KONTRADIKCIJE METODA

Usvojen je koncept visoko centralizovane digitalizacije. U stranoj literaturi odavno su prepoznati problemi koji se javljaju u tom kontekstu [1], [2]. Nije uočeno da se posebno analizirao problem pristupa podacima u kontekstu dualiteta digitalni-papirni dokument. Strategija i taktika su svedeni na pominjanje zakona o arhivskoj građi, obavezno navodjenje pojma „najbolje prakse“, i očekivanje brze digitalizacije kompletnog sistema državnih organa.

U početnoj fazi uvodjenja DMS-a neprekidno se govori o značaju mogućnosti pronalaženja bilo kog dokumenta, bilo gde, u državnom organu. U kasnijim fazama realizacije počinje se sa ograničavanjem pristupa saradnicima organa državne uprave. Uvode se ograničenja - filteri za dokumenta koje pojedine niže organizacione jedinice i pojedini saradnici mogu i treba da vide. Naknadna analiza ograničenja pristupa pokazuje da su umesto sveobuhvatnog „wide-range“ DMS-a, često mogli da se primene znatno jednostavniji (tj. jeftiniji) sistemi sa istom funkcionalnošću. Navodimo:

1. Svi isporučiooci DMS-a redovno ističu kao "glavnu" prednost korišćenja DMS-a mogućnost da svako vidi sve dokumente u DMS-u. Uopšte ne postoji potreba da svako vidi sve dokumente u DMS-u. Administrator DMS-a po funkciji može da ima pristup svim dokumentima, ali isključivo iz tehničkih razloga upravljanja repozitorijumima dokumenata.

2. Donosilac odluka suviše kasno shvata problem iz tačke 1. i onda uvodi restrikcije pristupa, koje su ustvari restrikcije na osnovne funkcije DMS-a. Cena pogrešne primarne odluke je previsoka.

3. Problem pronalaženja bilo kog dokumenta u klasičnom sistemu papirne dokumentacije, neposredno ukazuje na lošu

Nikola Popović, stalni sudski veštak za informacione tehnologije, e-mail: [nikola.popovic@eunet.rs](mailto:nikola.popovic@eunet.rs), [nikola.popovic@mfa.rs](mailto:nikola.popovic@mfa.rs), Beograd

Julijana Mirčevski, stalni sudski veštak za informaciono-komunikacionu tehnologiju, [julijana.mir@gmail.com](mailto:julijana.mir@gmail.com), Beograd

organizaciju posla u državnom organu. Nije zadatak DMS-a da koriguje lošu organizaciju posla. Dokument pogrešno dislociran u DMS je veoma teško pronaći (na primer: pogrešno imenovan fajl u pogrešnom folderu).

### III. SPECIFIČNOSTI DIGITALNIH DOKUMENATA

Ovde nije mesto za analizu razlike između strukturiranih i nestruktuiranih dokumenata. Dokumenti u domenu finansija i upravljanja kadrovima su u velikom procentu strukturirani. Dokumenti u ostalim domenima nisu strukturirani (note, dopisi, informacije, analize itd.). DMS softveri koji se u ovom trenutku uvode u državnoj administraciji po pravilu ne vode računa o ovoj razlici. Po našim saznanjima prodavci softvera insistiraju na iskustvima softverskog sistema SAP (Systems Applications and Products in Data Processing) koji zahteva obimno, suštinsko prilagodjavanje i unifikaciju poslovanja u posmatranoj organizacionoj celini. Državna administracija je po pravilu tematski organizovana (pravosuđe, finansije, zdravstvo, spoljni poslovi itd.). Praksa je pokazala da implementacija "univerzalnog" DMS-a po definiciji usmerenog na rad sa dokumentima kao baznim elementima, generiše nove, neočekivane probleme. Tematski orijentisana organizacija državnog organa zahteva značajno precizniju studiju izvodljivosti (eng. feasibility study). Prodavci DMS-a izbegavaju izradu studije stanja jer ima visoku cenu po vremenu a i zahteva iskusne ljudske resurse. Pokušaj zamene studije stanja anketom ne može da obezbedi pouzdan uvid u organizaciju rada u državnom organu niti da identifikuje bitne probleme.

Korisnički orijentisan pristup (eng. user-centric approach) u implementaciji DMS-a je u osnovi ispravan. Međutim, organizacija prostora dokumenata u državnoj administraciji i pojedinačnim agencijama ne može da bude korisnički orijentisana (eng. user-centric approach) već tematski orijentisana (eng. topic-centric approach). Insistiranje na zadovoljavanju potreba građana i uštedama u državnoj administraciji je jedan aspekt problema. Drugi a značajniji aspekt je da digitalizacija mora da obezbedi uslove za rešavanje potreba državnih organa kako bi oni stvarno mogli da budu efikasniji. Ovo je ključna protivrečnost koja se pojavljuje u usvojenom procesu digitalizacije u državnoj administraciji. Korisnički orijentisan pristup digitalizaciji i tematski orijentisan pristup digitalizaciji su suštinski različiti. Olakšavanje rada korisnika korišćenjem DMS-a ne treba da se svede na "štednju papira" i "smanjenje štetanja po hodnicima". DMS mora da korisniku obezbedi efikasno raspolaganje informacionim resursima državnog organa odnosno svake niže organizacione jedinice. Nužno je efikasno pretraživanje i mogućnost fleksibilnog, multiaspektnog povezivanja dokumenata. To je uslov za generisanje novih informacija iz raspoloživih dokumenata.

Praktično iskustvo pokazuje da prodavci ne vode računa o bitnim elementima:

- kvalitetu skeniranih dokumenata,

- koriste proizvod kako bi primarno smanjili svoje troškove organizacije implementacije i održavanja.

Isporučiocima ove vrste softvera tvrde da obavljaju OCR (Optical Character Recognition) proces na unetim dokumentima ali u nekim slučajevima je ustanovljeno da se

ne radi. Ovo je još jedna bitna protivrečnost digitalizacije u državnim organima. I dalje postoji promet papirnih dokumenata koje je, u postojećim uslovima, nemoguće koristiti samo u digitalnoj formi. Proces OCR-a generiše određen manji procenat grešaka u prepoznatom tekstualnom sadržaju dokumenta. U našoj državnoj administraciji uobičajeno se koristi font Times New Roman koji je značajno nepovoljan za proces OCR-a. Pored toga prodavci se ne upuštaju u kvalitet procesa digitalizacije papirnih dokumenata. Slike serijski skeniranih dokumenata po pravilu odstupaju od vertikalne osovine, što zahteva primenu automatske korekcije linearnosti redova teksta (eng. deskewing). Osvetljenost i kontrast slike se drži na prosečnim vrednostima. Prodavci tekstualni sadržaj dobijen OCR-om implementiraju neposredno u PDF dokument, bez provere ispravnosti teksta spelling checkerom. Pošto je tekst dobijen OCR-om ugrađen u PDF nije ga moguće ni vizuelno kontrolisati. Takvi digitalni dokumenti ne mogu da budu priznati za valjanu arhivsku gradju. Isporučiocima softverskih alata u sklopu DMS-a pažljivo čute o ovim problemima.

Metapodaci implementirani u PDF dokument su i dalje minimalni iako postoje efikasni softveri za generisanje self-contained dokumenata sa znatno većim brojem metapodataka.

U kancelarijskom poslovanju državne administracije uobičajeno je da se uz izlazni dokument dostavlja dopis koji sadrži opis sadržaja dokumenta, razlog za dostavljanje dokumenta, identifikaciju dokumenta itd. Sam dokument koji se prosledjuje je po pravilu prilog. U praksi dokument često ima više priloga. Prodavci uobičajeno povezuju jedan identifikacioni opis dokumenta sa jednim dokumentom (najčešće u PDF formatu u državnoj administraciji). Ovo već sada stvara značajne tehničke i administrativne probleme. Više priloga se pakuje u jedan PDF dokument čime se gubi struktura paketa dokumenata. Gube se veze između dokumenata ili se unose naknadni podaci o vezama. Problem je što tekuće izvedbe DMS sistema ne dozvoljavaju formiranje ad-hoc klastera dokumenata po jednoj ili više tema. U upotrebi je samo jedna unapred propisana klasifikacija, što je jasno sa stanovišta spoljašnjeg korisnika usluga državnog organa. Potreba za organizovanjem prostora dokumenata prema stvarnim, dinamičkim potrebama internog korisnika nije nigde sagedana. Ovakva koncepcija DMS-a bez ikakve potrebe smanjuje upotrebljivost repozitorijuma dokumenata. Iskustvo pokazuje da određeni open source programi ove opcije omogućavaju.

Korišćenje novih metoda izdvajanja ključnih reči, nezavisnih od jezika dokumenta i bez kontrolisanog rečnika, stvorilo je nešto veću mogućnost logičkog povezivanja dokumenata u procesu pretraživanja. Značaj se ogleda u tome što se ne zahteva rad profesionalnog indeksera dokumenata, a koji nisu ni predviđeni za rad u državnoj administraciji. Hijerarhijska klasifikacija dokumenata predviđena Uredbom o kancelarijskom poslovanju je ispod nivoa potreba rada sa digitalnim repozitorijumima dokumenata.

## IV. MOGUĆI PRAVCI PRIMENE DIGITALIZACIJE

Praktično ispitivanje mogućnosti aktuelnih softverskih paketa ukazalo je na moguće pravce razvoja. Logika dodeljivanja privilegija pristupa ukazala je da za segmente ograničene nadležnostima organizacionih jedinica može da se koristi Wordpress [6] u ulozi mikro DMS-a. Na primer arhivske fascikle mogu da se reprezentuju u formi Wordpress web sajta. Pri radu sa udaljenim filijalama (na primer diplomatsko konzularna predstavništva) mogu efikasno da koriste post opcije Wordpress-a, uz obezbeđenje sigurne komunikacije.

Uzroci nemogućnosti pronalaska nekog dokumenta u organu državne administracije prvenstveno leže u slaboj organizaciji rada. Pravila rada (pravilnici) ili nisu ažurirana, ili su loše definisana, ili se jednostavno ne poštuju. Primena glomaznih i veoma skupih DMS sistema sasvim sigurno ne može da reši navedene probleme.

Lista testiranih softverskih modula koji mogu da realizuju funkcionalnosti DMS u manjim organizacionim celinama:

**Wordpress - open source CMS software** [2] [6]

**nuBuilder - Rapid Application Developer**, sada ima priključak na Wordpress. Koristi relacionu baza podataka MySQL [6]

**Tematres** - softver za izradu tezaurusa i taksonomija i prikazivanje strukture repozitorijuma dokumenata [7]

**Tesseract free open source softver** za OCR (Optical Character Recognition) [8]

**ScreenReader** (komponenta programa ABBYY Fine Reader) koristi se za preuzimanje teksta sa ekrana

**YAKE** algoritam za ekstrakciju ključnih reči iz dokumenata (bez nadzora - eng. unsupervised method). Za nas je veoma važna karakteritika da rad algoritma ne zavisi od jezika algoritma. Koristi se u verziji programskog jezika **Python** bez dodatnog programiranja. [9] spelling checker za srpski jezik (bilo koja implementacija, preporučljivo je da se za rad koristi NoSQL verzija baze podataka na primer MySQL v.8 i dalje). Odredjena regulacija karakteristika može da se postigne izmenom liste tzv. stop reči.

**KNIME** (open-source data analytics, reporting and integration platform) izuzetno koristan "free machine learning" softver u domenu vizuelnog programiranja (topic aggregation, documents clustering for information consolidation, automatizing workflow itd.) [10]

Praktični eksperimenti ukazuju da već uz minimum programerskih kapaciteta (dva do tri programera srednjeg nivoa obučenosti za programske jezike PHP i Javascript), navedena grupacija open source softvera može da generiše prototipove DMS-a manjih dimenzija, ali bolje prilagodjenih potrebama državne administracije sa aspekta organizovanja prostora dokumenata. U stvari u pitanju su mikro DMS sistemi odnosno mikro CMS sistemi koji mogu da formiraju mrežu za razmenu podataka u slučaju potrebe.

Dosadašnje praktično iskustvo autora ukazuje da sveobuhvatni (eng. large scale) DMS sistemi koji se

trenutno implementiraju u državnoj administraciji ne ispunjavaju više bitnih uslova za pouzdanu organizaciju prostora dokumenata u državnoj administraciji.

Posebno konstatujemo da je istraživačka delatnost je praktično onemogućena. Izmene može da radi samo prodavac. Već na kratki rok pojaviće se ozbiljni problemi u funkcionalnosti implementiranih DMS-ova. Velike agregacije dokumenata neće biti fleksibilne i podložne reorganizaciji usled fiksne strukture DMS-ova. Uporno se ponavlja mantra "skalabilnosti", ali se nigde ne pominje da je u pitanju aspekt "fizičke skalabilnosti" (po fizičkim dimenzijama repozitorijuma dokumenata) a ne i "logičke skalabilnosti" (problem generisanja novih informacija, problem konsolidacije informacija vidi Saračević et all. [8]).

Kompanije proizvođači i isporučiooci ove vrste softverskih proizvoda redovno zaštićuju svoje proizvode autorskim pravima. Postavlja se pitanje zaštite vlasništva nad podacima i posebno mogućnosti upravljanja podacima od strane državnih organa. U ugovorima se problem migracije na drugi, noviji sistem ostavlja za "buduće vreme". Prodavci izbegavaju ozbiljan razgovor o temi migracije. Korišćenje usluga *Data centara* još uvek nije dovoljno precizirano, posebno u domenu softvera za upravljanje upravljanje podacima - šta se dešava sa softverom? Pitanje urgentne mobilnosti podataka i posebno DMS-a se retko postavlja iako nas bliska istorija (1999. godina) upozorava na veoma stvarne probleme čuvanja i očuvanja podataka državnih organa.

## V. DIGITALIZACIJA I SRPSKI JEZIK

Jezički aspekti digitalizacije su izostavljeni iz strategija digitalizacije. Information Retrieval problem se uopšte ne pominje u strategijama a neposredno je vezan za jezik i pismo, posebno u domenu državne uprave. Do sada istraživani metodi automatskog indeksiranja dokumenata nisu dali ni približno zadovoljavajuće - pozitivne rezultate. U korpusu dokumenata u domenu državne administracije koristi se jezik u formi koja najverovatnije nije pogodna za proces indeksiranja ("administrativni jezik"). Testiranja su obavljena na Open source softverima, ali veoma je velika verovatnoća da ni komercijalna rešenja ne mogu da daju bolji rezultat. Osnovni problem je što su postojeći jezički alati za srpski jezik veoma ograničenih mogućnosti ili ne postoje u potrebnoj formi (službeni rečnici pojmova, POS tageri, itd.). Na primer u programu Microsoft Word postoji modul Grammar koji izuzetno dobro koriguje pravopisne i gramatičke greške korisnika engleskog jezika i još nekoliko jezika koji su zastupljeni u svetu. Za srpski jezik ne postoji ništa slično čak ni u nagoveštaju. Nedostatak ključnih softverskih komponenti iz domena lingvistike i odgovarajućih jezičkih korpusa dovodi do toga da čitav sistem digitalizacije neće moći da se odvija onom dinamikom kojom se zamišlja. Vidljivost dokumenata i uopšte informacija, drastično je ograničena time što se kasni za tzv. jezičkom industrijom u svetu. Ovo je segment u kome su neophodne inovacije iz domena namenskog softvera, u veoma kratkom roku, a mladi istraživači se već pojavljuju [13] ali im je neophodna promptna podrška države.

## VI. ZAKLJUČAK

Uzevši u obzir dugogodišnje iskustvo u radu na parcijalnim procesima digitalizacije u državnim organima, savremeni trend razvoja informatike, tendenciju odvajanja softvera od hardvera, primenu metoda veštačke inteligencije u pretraživanju i druge tekuće prisutne pojave u razvoju informacionog društva, može se zaključiti da je neophodno poštovati sledeće bitne okvire:

- digitalizacija u okviru pojma održivog razvoja,
- digitalizacija treba da se implementira na *open source* softverskim platformama,
- digitalizaciju mora da prati razvoj informatičkog instrumentarijuma za domen srpskog jezika i pisma.

To zahteva angažovanje više specijalista za softver i hardver ali znatno olakšava rešavanje problema interoperabilnosti informatičke podrške tokom dužeg vremena i na različitim hardverskim platformama.

Digitalizacija u državnoj administraciji je složen proces. U radu su naznačene primedbe nastale na osnovu praktičnog iskustva i saznanja tokom uvođenja DMS-a u više institucija. Primedbe se odnose na procese koji generišu nove probleme. Očekivanja da će samo ishitrenom digitalizacijom u državnim organima da se ostvare uštede i ubrzanja rada nisu realna. Potvrđuje se ispravnost odavno poznatog pristupa evolucionog razvoja sistema. *Outsourcing* razvoja IT sistema i dalje ima visoku cenu.

Trend odvajanja softvera od hardvera, bezbednost podataka i otvorena softverska rešenja prilagodljiva rastućim potrebama korisnika, treba da ostanu postulati razvoja digitalizacije.

Nepoštovanje ove činjenice redovno dovodi do značajnih kašnjenja u implementaciji sistema i nekontrolisanog rasta cena projekata u celini.

## LITERATURA

- [1] Qu'est-ce que la documentation? Briet, Suzanne. Paris (1951).
- [2] <https://www.epa.gov/laws-regulations/>, posećeno 12.4.2022.
- [3] <https://www.srbija.gov.rs/dokument/45678/strategije.php> (Retrieved by 09.05.2020 22:24 GMT)
- [4] В. Масникоса: Израда структуре за заштиту преноса, чувања и приступа добро чуваним обавештењима (Système –Protect; Patent br.), персонална страница Вукашина Масникосе, <http://www.termoenergo.com/source/masnikosa/index.html>
- [5] EUROPEAN COMMISSION DIGITAL STRATEGY A digitally transformed, user-focused and data-driven Commission, EUROPEAN COMMISSION, Brussels, 21.11.2018 C(2018) 7118 final, [https://ec.europa.eu/info/sites/info/files/strategy/decision-making\\_process/documents/ec\\_digitalstrategy\\_en.pdf](https://ec.europa.eu/info/sites/info/files/strategy/decision-making_process/documents/ec_digitalstrategy_en.pdf), Retrieved by 24.05.2020.
- [6] <https://wordpress.org/> posećeno 21.4.2022.
- [7] <https://www.nubuilder.com/> posećeno februar 2022.
- [8] <https://vocabularyserver.com/web> posećeno januar-mart 2022.

- [9] <https://github.com/tesseract-ocr/tesseract> posećeno april 2022.
- [10] <https://github.com/LIAAD/yake> posećeno februar-april 2022.
- [11] <https://www.knime.com/> posećeno april 2022.
- [12] <https://unesdoc.unesco.org/ark:/48223/ptf0000047738> posećeno 5.3.2022.
- [13] Vuk Batanović, Nikola Ljubešić, Tanja Samardžić, SETimes.SR – A Reference Training Corpus of Serbian. Retrieved by 17.06.2020. [https://vukbatanovic.github.io/publication/jtdh\\_2018\\_sr/](https://vukbatanovic.github.io/publication/jtdh_2018_sr/)

Abstract - The current state of the digitalization process in Serbia is beginning to show the first, clear signs of contradictory phenomena. A significant number of document management systems are being introduced, especially in the state administration system. With obvious advantages, there are problems with the compatibility of new systems with the existing paperwork system. Harmonization of document availability and mode of operation is difficult. One part of the documents, in addition to the digital version, must still remain on paper. The costs of digitization are high, and these problems cause significant delays in implementation. The paper identifies possible technological solutions that can improve the sustainability of the digitization process.

Index Terms - digitalization; sustainability; methodology of digitalization; state administration.

### Current problems of digitalization in state administration

Nikola Popović, Julijana Mirčevski