

Primjena *Oracle Text*-a u sistemima za upravljanje elektronskim dokumentima

Studentski rad

Aleksandar Vidaković

Student drugog ciklusa studija

Univerzitet u Istočnom Sarajevu, Elektrotehnički fakultet

Istočno Sarajevo, Bosna i Hercegovina

aleksandar.vidakovic.m192@student.etf.ues.rs.ba

Sažetak — Pretraživanje elektronskih dokumenata je ključno da bi korisnik dobio pristup tačno onim dokumentima koji su mu u datom trenutku potrebni. Skupovi podataka i struktura dokumenata mogu biti proizvoljno kompleksni, a kriterijumi za njihovo pretraživanje različiti. Rad će opisati razvoj jednog *database-driven* rješenja za upravljanje elektronskim dokumentima. Poseban akcenat će biti na indeksiranju teksta pomoću *Oracle Text* tehnologije čiji će značaj i primjena biti kritički analizirani.

Ključne riječi – dokumenti; indeksiranje; pretraga; *Oracle Text*.

I. UVOD

Dokument predstavlja strukturiranu jedinicu ili skup informacija koji može postojati u fizičkom (papirnom) ili elektronskom obliku. Često se koriste kao osnovna forma poslovne korendencije, dokazivanja (u pravnom, poslovnom i administrativnom okruženju), evidentiranja (dogadaja, akcija, transakcija) i formalizacije poslovnih procesa. U neke primjere poslovnih dokumenata se ubrajaju: računi, fakture, ugovori, pravilnici, procedure, planovi, zapisnici sa sastanaka i slično. Skup međusobno povezanih dokumenata koji se odnose na jedinstveni projekat ili poslovni predmet ili proces unutar organizacije jeste dokumentacija. U kontekstu poslovanja organizacija, postojanje dokumentacije je veoma važno radi usaglašenosti sa standardima i zakonskim regulativama, interne kontrole i izvještavanja i osiguravanja kontinuiteta kroz arhivu. Još jedan važan termin u sklopu ovog domena jeste poslovni proces koji se definiše kao skup aktivnosti koje je potrebno sprovesti da bi se stvorila određena vrijednost za korisnika ili organizaciju kroz transformaciju prosljeđenih ulaznih u očekivane izlazne podatke (npr. prvo se podnosi dokument zahtjeva za odobrenje odlaska na poslovni put, zatim se dobija i dokument samog odobrenja, nakon toga se predaje izvještaj sa puta i konačno se dobija dokument refundacije troškova) [1].

Sada kada je jasno zašto je dokumentovanje kritično za funkcionisanje organizacija i kompanija, potrebno je objasniti tradicionalne načine vođenja dokumentacija te istaći nedostatke istih iz čega se nedvosmisleno javlja poslovna potreba za softverskim rješenjem. Prvobitno su, u poslovnim okruženjima, dokumenti vođeni fizičkim čuvanjem u papirnom obliku u velikim arhivama, registratorima, ormanima ili radnim stolovima. Ovo je predstavljalo značajna izdavanja za

organizaciju (troškove za kupovinu resursa pa čak i iznajmljivanje zasebnog prostora) gdje su se dokumenti sa teškoćama nalazili što je u velikoj mjeri usporavalo poslovne procese. Iako pretraživanje može biti olakšano rasporedom dokumenata prema nekom kriterijumu (prema tipu, autoru ili datumu), zaposlenima je potrebno mnogo vremena za traženje kroz fascikle čime se smanjuje produktivnost, a povećava frustracija. Poslovni procesi mogu trajati danima dok se dokumenti ručno obrađuju putujući od kancelarije do kancelarije. Razmjena dokumenata vršena je manuelno ili klasičnom poštom.

Sa druge strane, vremenom i razvojem računarskih tehnologija počelo se sa čuvanjem dokumenata u *Word*, *Excel*, *PDF* ili drugim formatima u folderima na lokalnim računarima ili u okviru mrežno dijeljenih foldera. Međutim, tu nije omogućeno verzionisanje fajlova, automatizacija u smislu obavještavanja i izvještavanja, niti praćenje aktivnosti i kolaboracije. Ni kontrola pristupa nije obezbjeđena na aplikativnom nivou što može dovesti do curenja podataka i narušavanja reputacije.

Zbog gore navedenih problema, organizacije prepoznaju da je potrebno da krenu putem digitalne transformacije što podrazumijeva i automatizaciju rada sa dokumentima. Razlog za to je što klasični načini vođenja dokumenata nisu adekvatni da zadovolje potrebe tržišta gdje broj dokumenata sve više raste čime se kontrola nad istima značajno otežava. Potrebno je eliminisati jednostavne, ali ponavljajuće zadatke koje zaposleni rade i primijeniti dostupne tehnologije povećavajući time produktivnost i smanjujući operativne troškove [1]. U tom kontekstu, sistem za upravljanje elektronskim dokumentima (engl. *Document Management System - DMS*) se javlja kao logično rješenje koje omogućava:

1. Centralizovano čuvanje dokumenata,
2. Kontrolu pristupa po korisnicima i korisničkim ulogama,
3. Verzionisanje dokumenata i praćenje aktivnosti,
4. Automatsko indeksiranje i pretragu dokumenata,
5. Kolaboraciju – diskusije i zajednički rad na dokumentu,
6. Obavještavanje korisnika preko SMS-a ili *mail*-a o akcijama izvršenim nad dokumentom.

II. SISTEMI ZA UPRAVLJAJE ELEKTRONSKIM DOKUMENTIMA

A. Pojam, značaj i standardi

DMS je softversko rješenje koje preduzećima olakšava da organizuju, upravljaju i prate elektronske dokumente. To je centralizovana platforma gdje korisnici mogu kreirati, uređivati, skladištiti i brzo pretraživati dokumente. Upravljanje dokumentima pomoću DMS-a predstavlja primjenu računarskog sistema i softvera za čuvanje, vođenje i praćenje elektronskih dokumenata u šta spada: hvatanje sadržaja (engl. *capture*), automatizacija poslovnih tokova (engl. *workflow automatization*), skladištenje dokumenata, arhiviranje, izvještavanje, pretraga i pronalaženje konkretnih informacija te procesi za praćenje i kontrolu dokumenata. DMS sistemi sve više postaju dio širih sistema za administraciju digitalnog sadržaja (engl. *Enterprise Content Management – ECM*) s obzirom da podaci (digitalni sadržaj) pristižu u raznim struktuiranim i nestruktuiranim formatima (zvuk, tekst, slika, video) i sa raznih eksternih izvora kao što su internet, mobilni telefoni, USB i drugi [2].

ISO 12651-2 definiše dokument kao zabilježenu informaciju koja postoji kao samostalna cjelina, dok ISO 15489 uspostavlja principe upravljanja zapisima kao poslovnim dokazima i informacionom imovinom, uz obaveznu ulogu metapodataka u njihovoj kontekstualizaciji. Kvalitetno upravljanje elektronskim zapisima povećava transparentnost, odgovornost i usklađenost sa propisima, smanjuje rizike i troškove, štiti prava organizacije i korisnika, te obezbjeđuje kontinuitet poslovanja i podršku pravnim i operativnim procesima [3]. Dodatni standardi koji proširuju regulativni okvir su: ISO 30300 koji uvodi sistemski pristup i sertifikaciju upravljanja zapisima, ISO 23081 koji definiše metapodatke i interoperabilnost, ISO 19005 (PDF/A) koji obezbjeđuje dugoročno očuvanje digitalnih dokumenata, dok ISO 16175 određuje funkcionalne zahtjeve za moderne DMS aplikacije.

Prema [4], kompanije koje koriste DMS bilježe do 30% veću produktivnost zahvaljujući bržem radnom toku i pretraživanju dokumenata, a globalno tržište ovih rješenja bi do 2027. moglo dostići 11,06 milijardi dolara. Čak 70% organizacija smatra DMS ključnim za digitalnu transformaciju, dok se vrijeme pretrage dokumenata smanjuje za oko 50%. Više od polovine DMS rješenja danas je dostupno u oblaku, što omogućava rad na daljinu i lakši pristup dokumentima. Primjenom DMS-a organizacije ostvaruju značajne finansijske uštede (oko 30%).

Prema blogu *TheECMConsultant*, ključni izazovi u implementaciji DMS sistema u 2025. godini su se odnosili na rastuću složenost pretrage i obrade dokumenata, tehničke integracije sa postojećim sistemima, sigurnosne i zahtjeve za skalabilnošću, te otpor korisnika pri usvajanju novih rješenja. Najveća pažnja usmjerena je na efikasno upravljanje metapodacima, smanjenje rizika od gubitka dokumenata i osiguranje dostupnosti kroz *cloud* platforme. Kao odgovor na ove izazove, preporučuje se usvajanje strateškog pristupa upravljanju dokumentima i navode se korisne prakse koje omogućavaju veću sigurnost, bržu pretragu i stabilan rast DMS okruženja u skladu sa savremenim poslovnim potrebama [5].

B. Funkcionalnosti

U nastavku će biti nabrojane i ukratko opisane najvažnije funkcionalne cjeline koje DMS treba da ispuni.

Unos dokumenata predstavlja početnu i jednu od ključnih faza u DMS-u, jer određuje kvalitet kasnije pretrage, dostupnosti i obrade. Dokumenti se u sistem uvode direktnim korisničkim unosom, automatski kroz integracije sa drugim aplikacijama (putem REST/SOAP servisa, PL/SQL procedura ili *middleware* rješenja), skeniranjem papira u nadzirane mrežne foldere uz opcione OCR mehanizme ili kreiranjem na osnovu unaprijed definisanih šablona radi standardizacije. Metapodaci ovdje imaju centralnu ulogu jer opisuju dokument, određuju kriterijume pretrage, status, pravo pristupa i verziju, te prolaze validacionu logiku kako bi se spriječile greške i osigurala ispravna obrada.

Bezbjednost u sistemima kao što je DMS zasniva se na zaštiti dokumenata od neovlašćenog pristupa kroz kontrolu korisničkih privilegija, u skladu sa ISO 27001 standardom. Ključna mjera je autorizacija po principu minimalnih privilegija. U praksi se koriste različiti modeli kontrole pristupa: RBAC (dodjela prava kroz uloge što je skalabilno i jednostavno), ABAC (dinamičko odlučivanje na osnovu atributa korisnika, dokumenta i konteksta), MAC (najstroži model sa centralizovanim pravilima i nivoima tajnosti) i DAC (vlasnik dokumenta određuje pristupna prava).

Kada je dokument unesen i zaštićen pristupnim pravima, počinje kolaborativni rad više korisnika na istom sadržaju. DMS omogućava timsku obradu uz praćenje istorije izmjena, komentara i integraciju sa drugim poslovnim platformama. Kolaboracija može biti sinhrona (istovremeni rad u realnom vremenu) ili asinhrona (rad u različito vrijeme uz mehanizme zaključavanja i rješavanje konflikata). Verzionisanje je ključno za čuvanje verzija dokumenta, identifikovanih brojem, vremenom i autorom.

DMS treba da obezbjedi obavještenja koja se pokreću događajima (kreiranje, izmjena, brisanje dokumenta). Ona mogu biti *e-mail*, *push* u aplikaciji, mobilne notifikacije ili poruke kroz integracije (npr. sa *Slack*, *MS Teams*). Precizno definisanje kada i kome se šalju sprječava preopterećenje informacijama i podržava efikasan timski rad.

Skladištenje dokumenata bira se prema prioritetima sistema. Tip podataka BLOB u bazi osigurava sigurnost, ACID konzistentnost i kontrolu pristupa, ali usporava rad s velikim fajlovima. Fajl sistem čuva dokumente van baze (u bazi samo putanja), što donosi bolje performanse i skalabilnost, ali ne i transakcionu povezanost i potpunu dosljednost. Objektno skladištenje nudi maksimalnu skalabilnost, integracije i verzionisanje, uz moguće kašnjenje i slabiju konkurentnost. Hijerarhijsko skladištenje automatski dijeli dokumente po slojevima memorije radi optimizacije, ali je tehnički složenije za podatke koji se rjeđe koriste.

Arhiviranje znači premještanje neaktivnih, ali vrijednih dokumenata u zaštićen, dugoročni prostor pohrane uz poštovanje definisanih pravila čuvanja, pristupa i brisanja. Proces uključuje identifikaciju dokumenata za arhivu, dodjelu roka čuvanja, eventualnu konverziju formata, dodavanje metapodataka i fizičko premještanje u sigurnu zonu.

C. Pretraživanje dokumenata

Pretraživanje dokumenata je proces pronalaznja dokumenata prema korisnički definisanim kriterijumima, poput metapodataka, ključnih riječi ili fraza. Indeksiranje je sistemska obrada i klasifikacija dokumenata sa dodjelom metapodataka radi efikasnog pronalaznja u velikim skupovima podataka. Benefiti uključuju veću produktivnost, lakše ispunjavanje regulatornih zahtjeva, brže donošenje odluka, bolju timsku saradnju i smanjenje dupliranih podataka.

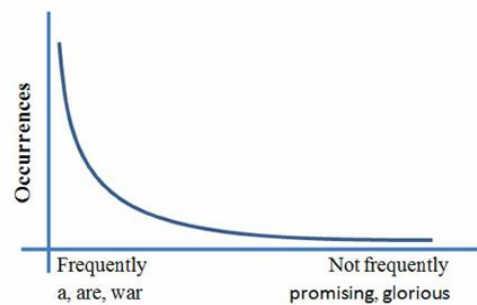
Neke od metode indeksiranja uključuju [6]:

1. Indeksiranje metapodataka: dokumenti se organizuju i pretražuju prema unaprijed definisanim atributima, poput autora, datuma ili tipa dokumenta. Prednost je manja potrošnja memorije i brži upiti, dok je ograničenje što se pretraga zasniva samo na metapodacima pa nedostaje fleksibilnost.
2. Indeksiranje oznakama: dokumentima se dodjeljuju korisnički definisani tagovi za slobodnu i fleksibilnu kategorizaciju. Prednost je agilnost i jednostavno grupisanje prema više kriterijuma, dok nedostatak predstavlja slabija kontrola i moguća nekonzistentnost tagova.
3. Hibridno indeksiranje: kombinuje metapodatke i tagove radi veće preciznosti i relevantnosti pretrage. Pogodno je za heterogene dokumente u kompleksnim organizacijama, ali povećava tehničku složenost i opterećenje sistema.
4. Automatsko indeksiranje: softver sam analizira i dodjeljuje metapodatke, često koristeći AI i mašinsko učenje. Omogućava visoku brzinu, skalabilnost i smanjuje ljudske greške, ali zahtjeva složenu implementaciju i obuku modela.

U skladu sa temom rada, detaljnije će biti opisana metoda indeksiranja cijelog teksta (engl. *full-text indexing*) koja omogućava indeksiranje svake pojedinačne riječi koja je sadržana u dokumentu čime kompletan sadržaj dokumenta (u opštem slučaju nestruktuiran) postaje pretraživa baza. Pretraživanje se vrši prema pojmu ili frazi kao grupi smisleno formulisanih pojmova. Primjenjuju se algoritmi za obradu teksta po principima: tokenizacije (razbijanja teksta na riječi ili fraze), redukovanje riječi na korijen, izostavljanje često korišćenih riječi (veznici, prijedlozi) i rangiranje rezultata analize teksta prema relevantnosti. Rezultat je unaprijed pripremljeni indeks kojim su podaci iz sadržaja dokumenta i sam dokument jednoznačno mapirani. Može se koristiti u digitalnim bibliotekama, pravnim i istraživačkim bazama. Ova metoda je korisna kada se radi o neuređenim tekstovima, fleksibilna je jer kontekstualnošću proširuje pretragu mimo informacija iz metapodataka i podržava komplekse upite. Međutim, ovakvi indeksi zauzimaju mnogo memorijskog prostora jer su višestruko veći od samih dokumenata. Neophodna je kvalitetna i moćna serverska ili *cloud* infrastruktura, a implementacija može biti kompleksna posebno kada je u pitanju lokalizacija (neki jezici imaju jako komplikovanu morfologiju). Tehnologije i alati koje realizuju ovu metodu su: *Elasticsearch*, *Oracle Text*, *PostgreSQL Full-Text Search*, *Apache Lucene* [6].

Pretraga cijelog teksta omogućava pretraživanje dokumenata i baza podataka prema sadržaju, što daje preciznije i specifičnije rezultate. Korišćenje tekstualnih indeksa je efikasnije od tradicionalnih metoda pretrage jer se pretraga zasniva na terminima, a ne na obrascima, i podržava različite tipove pretrage, uključujući pretragu termina i fraza, prefiksa, blizine, operatore istinitosti, *fuzzy* pretragu i pretragu preko pogleda baze podataka. Glavni izazovi ove metode obuhvataju skalabilnost, relevantnost rezultata, rukovanje sinonimima, skraćenicama, homonimima, fonetikom, greškama u kucanju, pretragom slika i reindexiranjem dinamičkih podataka [7].

Upit za pretragu se sastoji od nekoliko riječi, a rezultat pretrage je lista dokumenata koji sadrže tu riječ. Za svaku riječ u svakom dokumentu potrebno je sačuvati zapis u indeksu. Kao značajan parametar, pominje se udaljenost termina u dokumentu (engl. *proximity*) sa idejom da su termini koji se pojavljuju bliže jedan drugom semantički povezani pa su i rezultati pretrage relevantniji. Takođe, važna je i učestalost termina u dokumentu što određuje maksimalno vrijeme odziva upita (ilustracija i primjeri na Sl. 1) [8].



Sl. 1 Frekvencija pojmova u dokumentu [8].

III. ORACLE TEXT

Oracle Text je moćna funkcionalnost u okviru *Oracle* baze podataka koja omogućava kreiranje, upravljanje i korišćenje pretrage cijelog teksta nad tekstualnim podacima, dokumentima i tekstualnim atributima u tabelama. Glavna prednost je kompatibilnost sa *Oracle SQL*-om čime je moguća direktna integracija u sloj baze podataka.

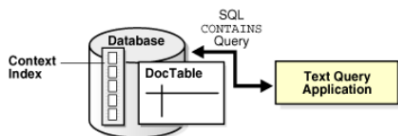
Centralna komponenta *Oracle Text*-a je domenski indeks (engl. *domain index*) koji je specijalizovan za obradu teksta. Nakon što se indeks kreira klasičnom naredbom *create index*, sistem može izvršavati upite za tekstualnu pretragu. U sljedećoj tabeli (Tablea I) su navedeni neki PL/SQL paketi koji služe za rad sa *Oracle Text* indeksima.

TABELA I. PL/SQL PAKETI ZA RAD SA ORACLE TEXT INDEKSIMA.

| PLSQL paket | Namjena |
|-------------------|--|
| <i>ctx_ddl</i> | Kreiranje, brisanje i održavanje indeksa. |
| <i>ctx_query</i> | Analiziranje i optimizacija upita. |
| <i>ctx_report</i> | Generisanje izvještaja o indeksiranim dokumentima. |
| <i>ctx_doc</i> | Rad sa pojedinim dokumentima u indeksu. |
| <i>ctx_user</i> | Provjera korisničkih prava nad indeksima. |

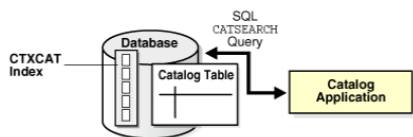
Neki od najčešćih tipova *Oracle Text* indeksa su [9]:

1. CONTEXT koji je pogodan za pretragu velikih količina tekstualnog sadržaja raznih nestruktuiranih formata sa naprednim mogućnostima uz podršku za leksičke preference i filtriranje. Internim mehanizmima vrši tokenizaciju teksta čuvajući bogat skup metapodataka (tip tokena, ID dokumenata u kojima se prvi i posljednji put pojave token i BLOB sa informacijama o frekvenciji i poziciji tokena unutra teksta). Može koristiti funkciju *contains* kao što je ilustrovano na Sl. 2;



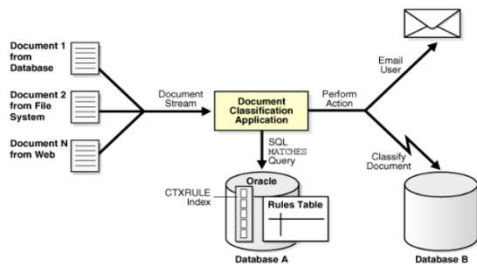
Sl. 2 Context indeks [9].

2. CTXCAT koji je namjenjen za brzu pretragu strukturiranih skupova podataka manjeg i srednjeg obima (Sl. 3). Optimizovan je za upite sa *where* i *order by* klauzulama nad kolonama sa indeksiranim tekstom. Može koristiti funkciju *catsearch*;



Sl. 3 Ctxcat indeks [9].

3. CTXRULE koji služi za klasifikaciju i automatsko prepoznavanje dokumenata prema unaprijed definisanim pravilima (Sl. 4). Na primjer, može koristiti upite sa *matches* funkcijom.



Sl. 4 Ctxrule indeks [9].

U upitima koji se šalju bazi sa uslovima nad kolonama koje su indeksirane *Oracle Text*-om koriste se brojni operatori. Klasični logički operatori upravljaju logičkim odnosima između pojmova. Sa druge strane, operatori kao što su *near*, *whitin* ili *about* proširuju kontekst pretrage uvodeći konceptualne veze. I dodatno, radi povećanja fleksibilnosti i tolerancije sistema na razlike u pisanju, često se koriste *fuzzy*, *stem* ili *wildcard* operatori [9].

U pozadini *Oracle Text* indeksa leži interni mehanizam koji čine indeksno organizovane tabele koje se automatski generišu nakon kreiranja tekstualnog indeksa sadržavajući podatke zavisno od tipa indeksa. Naziv im je u formi *DR\$<indeks>\$< sufiks>* gdje je *<indeks>* ime tekstualnog indeksa, a *< sufiks>* oznaka tipa odgovarajuće indeksno organizovane tabele koja podržava indeks [9].

IV. PRAKTIČNA REALIZACIJA

Razvijena je troslojna veb DMS aplikacija kao dio šireg informacionog sistema primjenom *Oracle* tehnologija (*Oracle RDBMS*, *PL/SQL*, *APEX*) čija je arhitektura data na Sl. 5. Krajnji korisnici pomoću veb čitača pristupaju prezentacionom sloju. HTTP zahtjevi se konvertuju u konkretne pozive *PL/SQL* funkcionalnosti iz baze posredstvom srednjeg sloja u vidu *APEX Engine*-a sa komponentom *Oracle REST Data Service – ORDS*.

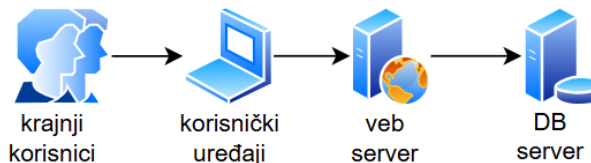
U modelu baze podataka primarni entiteti su podijeljeni u dvije glavne grupe. Prva grupa se odnosi na upravljanje dokumentima i uključuje podatke o kategorijama, tipovima, lokacijama dokumenata, registratorima, korisničkim folderima, starim verzijama dokumenata te tabele koje povezuju korisnike i radne pozicije sa pravom pristupa određenim dokumentima. Drugu grupu čini skup tabela za upravljanje poslovnim predmetima. Te tabele su: klasifikacije, statusi i način sprovođenja predmeta, šabloni poslovnih predmeta i pratećih dokumenata kao i konkretne instance poslovnih predmeta i njihova dokumentacija.

U implementiranom DMS-u dokumenti se unose na tri načina zavisno od izvora podataka. Prvi je klasičan *upload* kroz *APEX* formu gdje se fajl privremeno smješta u *apex application temp tables*, a zatim u tabelu upisuje kao *BLOB* tip podataka automatskim dodjeljivanjem metapodataka. Drugi način je uvoz iz dijeljenog foldera gdje se fajlovi skladište nakon skeniranja, a potom ubacuju u DMS pomoću *Java* uskladištenih procedura za rad sa fajl sistemom te biblioteka *dbms_lob* i *utl_file*. Treći način je automatsko kreiranje poslovnih predmeta sa pripadajućim priložima iz drugih modula preko interfejsa definisanih *PL/SQL* paketima.

Pri kreiranju dokumenta korisnik odmah definiše pristupne privilegije biranjem vidljivosti i tipa dokumenta, čime se automatski učitavaju inicijalna pravila pristupa. Zatim, sistem vodi korisnika na stranicu za upravljanje privilegijama gdje su prikazane liste sa podjelom po korisnicima ili radnim pozicijama. Pravila za čitanje i izmjenu se mogu prilagođavati. Drugi nivo kontrole je realizovan *APEX* šemama autorizacije radi regulisanja vidljivosti elemenata.

Omogućeno je kolaborativno uređivanje dokumenata tako što se svaka izmjena evidentira kroz nove verzije sa metapodacima, pri čemu se trenutna verzija čuva u glavnoj tabeli, a prethodne se putem trigeru arhiviraju u zasebnu tabelu. Nova verzija se može dodati klasičnim *upload*-om ili naprednije, korišćenjem *APEX Office Edit (AOE) plugin*-a koji omogućava direktno uređivanje u veb čitaču.

Obavješavanje privilegovanih korisnika o izmjenama automatizovano je kroz slanje mejla i evidentiranje poslatih poruka u bazi. Dodatnu dimenziju kolaboracije čine diskusije o dokumentima, gdje se vode tematski razgovori korisnika.



Sl. 5 Arhitektura rješenja.

digitalizacije dokumentacije. U radu je prvenstveno izvršeno teoretsko razmatranje DMS sistema od same definicije ključnih pojmova preko standardizovanih regulativa iz ove oblasti i najboljih savremenih praksi do analize ključnih koncepata koje moderni DMS sistemi moraju podržavati.

U skladu sa naslovom rada, posebna pažnja je posvećena problemu pretraživanja dokumenata u DMS sistemima gdje je istaknut značaj izbora odgovarajuće strategije indeksiranja podataka unutar baze. Nakon što su navedene moguće opcije, detaljnije je opisana metoda pretraživanja samog sadržaja dokumenta koja je najviše odgovarala kontekstu kreiranog sistema.

Praktični dio rada ukratko opisuje razvoj jednog *database-driven* DMS rješenja u vidu veb aplikacije koja je poslužila za demonstraciju pomenutih koncepata. Obrada unosa dokumenata je u potpunosti prilagođena poslovnim potrebama podržavajući različite izvore podataka. Za kontrolu pristupa sistem kombinuje opšte šeme autorizacije, ali i specifične korisnički definisane privilegije čime se postiže fleksibilnost prema poslovnim potrebama. Kolaborativni rad je obezbjedio transparentno uređivanje dokumenata uz slanje obavještenja, praćenje istorije izmjena i efikasnu komunikaciju među korisnicima.

Kritično važno je bilo realizovati što intuitivniju pretragu dokumenata. Radi kompatibilnosti sa drugim tehnologijama koje su korišćene izabrana je primjena *Oracle Text*-a za kreiranje domenskih (tekstualnih) indeksa. Radi se o moćnom, ali zahtjevnom rješenju. Izazovi u primjeni se odnose na kompleksnost konfiguracije i dodatnih prilagođavanja. Pristupna je i potencijalna zavisnost od DBA ekspertiza radi budućih optimizacija u odnosu na neke druge *cloud* pretraživačke servise. Međutim, prednost je duboka integracija sa bazom podataka i širok spektar naprednih mogućnosti indeksiranja. Konkretno rješenje je dizajnirano kao jedan od modula postojećeg ERP sistema što je diktirano i sami izbor tehnologija. Krajnjim korisnicima je omogućeno da dokumente pretražuju lako i jednostavno pisanjem prirodnog teksta koji se analizira formirajući validan upit ka bazi.

ZAHVALNICA

Rad je nastao u okviru predmeta Upravljanje elektronskim dokumentima pod mentorstvom doc. dr Snježane Milinković kojoj se autor zahvaljuje na savjetima.

LITERATURA

- [1] S. Jordan, S. S. Zabukovšek, I. Š. Klančik, (Naše gospodarstvo/Our economy), juni 2022, "Document Management System – A way to digital transformation", [Online] dostupno na: https://www.researchgate.net/publication/361922600_Document_Management_System_-_A_Way_to_Digital_Transformation.
- [2] Assosiation for Intelligent Information Management, "What is Document Management System (DMS)?", [Online] dostupno na: <https://www.aiim.org/what-is-document-imaging>.
- [3] International Organization for Standardization, "ISO 15489 - Information and documentation - Records management - Part 1: Concepts and principles", [Online] dostupno na: <https://www.iso.org/obp/ui/en/#iso:std:iso:15489:-1:ed-2:v1:en>.
- [4] Artsyltech, "Document Management Systems (DMS): Types, Uses, Technologies", [Online] dostupno na: <https://www.artsyltech.com/Document-Management-Systems-DMS#what-is-the-value-of-a-document-management-system>.
- [5] The ECM Consultant, "11 Document Management Best practices & Tips", [Online] dostupno na: <https://theecmconsultant.com/document-management-best-practices/>.
- [6] Arya.ai, "Learn Everything You need to know about Document Indexing", [Online] dostupno na: <https://arya.ai/blog/document-indexing>.
- [7] C. Singh, V. Gupta, "A Review of Full-Text Search in Different Perspective", Proceedings of the 2nd ICSSR Conference on "India Towards Viksit Bharat @2047", 13th– 14th September, 2024 Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA), 2024, [Online] dostupno na: <https://bvicam.in/INDIACom/news/ViksitBharat2024Proceedings/papers/38.pdf>.
- [8] A. B. Veretennikov, "Proximity Full-Text Search with a Response Time Guarantee by Means of Additional Indexes", Intelligent Systems Conference 2018, septembar 2018, [Online] dostupno na: <https://arxiv.org/pdf/1811.07361>.
- [9] *Oracle Documentation, Oracle Text Application Developer's Guide*, [Online] dostupno na: <https://docs.oracle.com/en/database/oracle/oracle-database/21/ccapp/understanding-oracle-text-application-development.html>.

ABSTRACT

Searching electronic documents is essential for providing users with access to exactly the information they need at a given moment. The data sets and document structures can be arbitrarily complex, and the criteria for searching them may vary. This paper will describe the development of a database-driven solution for electronic document management. Special emphasis will be placed on text indexing using Oracle Text technology, with its significance and application being critically analyzed.

THE APPLICATION OF ORACLE TEXT IN ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS

Aleksandar Vidaković