

Uvod u Tehnologije i platforme za upravljanje elektronskim sadržajima i dokumentima

Dragan Ivanović
dragan.ivanovic@uns.ac.rs

Katedra za informatiku, Fakultet tehničkih nauka, Novi Sad

2024.

Ko će držati nastavu

- Dragan Ivanović, NTP 329
dragan.ivanovic@uns.ac.rs
- Ivan Mršulja, NTP 322
ivan.mrsulja@uns.ac.rs

Oblasti kojima ćemo se baviti

- Upravljanje dokumentima
document management
- Pronalaženje informacija
information retrieval
 - Modeli
 - Performanse

Osnovna literatura

- D. Ivanović, B. Milosavljević. *Upravljanje digitalnim dokumentima*. Fakultet tehničkih nauka, 2015.
- C.D. Manning, P. Raghavan, H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- R. Baeza-Yates, B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, 1999.
- Apache Lucene, <http://lucene.apache.org>
- O. Gospodnetić, E. Hatcher. *Lucene In Action, Second edition*, Manning, 2010.

Program rada

Redni broj	Tema
1	Uvod u Tehnologije i platforme za upravljanje dokumentima
2	Sistemi za upravljanje digitalnim dokumentima i standardizacija
3	Pretraga tekstualnih dokumenata
4	Modeli - Bulov model
5	Modeli - Vektorski model
6	Performanse sistema za pretraživanje
7	Primeri projekata: CRIS, eSed, ENGAGE
8	Prezentacija projektnog zadatka

Program rada

Nedelja	Vežbe
1	—
2	—
3	Uvod i indeksiranje
4	Pretraživanje
5	Analiza srpskog sadržaja
6	Obrada PDF dokumenata
7	Spring boot i Elasticsearch
8-14	Izrada projekta

Kako se polaže ispit

- Polaže se teorija pismeno/usmeno
- Odbrani se samostalan projekat

Pojam dokumenta

- Pojam dokumenta obuhvata
 - tradicionalne papirne dokumente
 - računarski obrađene informacije kojima se **rukuje kao osnovnom jedinicom obrade**

Digitalni dokumenti

- Digitalno doba
- Razvoj ICT
 - Razvoj WWW
 - Razvoj hardvera
 - The man with the perfect memory - Gordon Bell, 1TB dovoljan za 83 godine života bez videa, 200TB memorije uključujući video
- Primeri:
 - tekstualni dokumenti, npr. tekstualni opisi ili poruke
 - grafički dokumenti, npr. slike, crteži, dijagrami, grafikoni
 - struktuirani dokumenti, npr. HTML i XML+XLink dokumenti
 - mediji sa vremenskom dimenzijom: zvuk, video
 - kompozitni multimedijalni dokumenti: sastavljeni od teksta, slike, zvuka, ili videa
 - višejezični dokumenti

Šta je metapodatak?

- Metapodaci = podaci o podacima
- Sadržaj elektronskog dokumenta predstavlja podatke
- Podaci o dokumentu predstavljaju metapodatke
- Postoje li metapodaci za metapodatke?
 - formalno mogu da postoje
 - pitanje je koliko to ima smisla

Primeri

- Primer 1: metapodaci za tekstualni dokument
 - autor
 - naslov
 - datum nastanka
 - ključne reči
- Primer 2: metapodaci za fotografiju
 - autor
 - datum i vreme fotografisanja
 - mesto fotografisanja
 - podešavanje aparata
 - objekti prikazani na slici

Svrha metapodataka

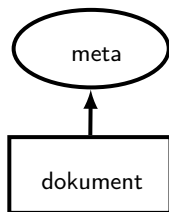
- Najvažnija svrha metapodataka
 - organizovanje kolekcije sadržaja (dokumenata)
 - klasifikacija dokumenata
 - pretraživanje dokumenata

Izvori metapodataka o dokumentu

- Životni ciklus dokumenta
- Poslovni proces gde se dokument koristi kao nosilac informacija između aktivnosti
- Opšta baza znanja u organizaciji u kojoj se odvija poslovni proces

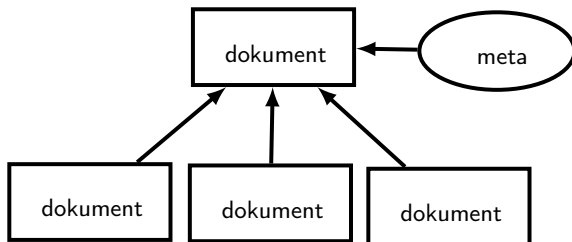
Pojedinačni dokument

- Elementarni oblik nosioca informacija
- Ima pridružene **metapodatke** koji opisuju njegov sadržaj ili druge karakteristike
 - metapodaci: podaci o podacima



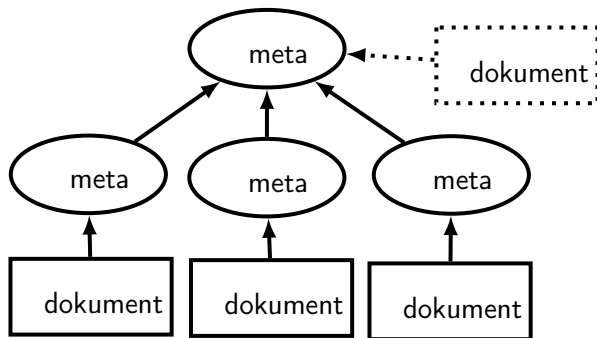
Složeni dokument

- Kompozicija više dokumenata različitih tipova
 - npr. tehnička specifikacija koja se sastoji od tekstualnih fragmenata, crteža, i dijagrama
- Metapodaci se pridružuju složenom dokumentu kao celini



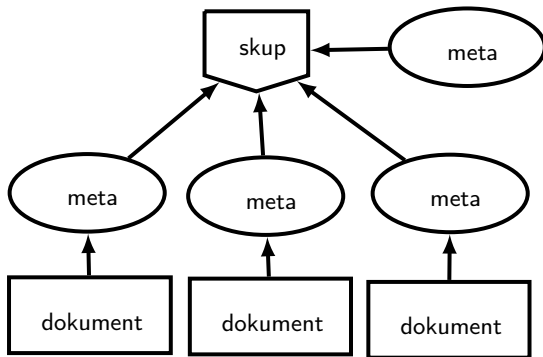
Agregacija dokumenata

- Skup samostalnih dokumenata, svakog sa svojim metapodacima
- Agregacija poseduje sopstvene metapodatke
- Može, a ne mora, da poseduje poseban sopstveni dokument



Skup dokumenata

- Posедује sopstvene metapodatke
- Svrha skupa, kao i sadržanih dokumenata, opisana je metapodacima



Veze između dokumenata

- **Aktivna veza:** stanje u kome deo sadržaja jednog dokumenta biva preuzet ili na neki drugi način zavisi od sadržaja drugog dokumenta
 - izmenom drugog dokumenta menja se i prvi
- Šta ako se promeni drugi dokument? Da li promena treba da bude vidljiva i u prvom?
 - ...upravljanje verzijama dokumenta

Životni ciklus dokumenta

- Dokument menja stanje u toku svog postojanja
- Upravljanje dokumentima = pravila i procedure za rukovanje dokumentima u toku životnog ciklusa
- Faze životnog ciklusa:
 - inicijalizacija
 - priprema
 - uspostavljanje
 - korišćenje
 - revizija
 - arhiviranje
 - uklanjanje

Inicijalizacija dokumenta

- Formiranje podataka potrebnih za kasniju pripremu
- Ne obuhvata pripremu i utvrđivanje sadržaja
- Rezultat je okvir u kome se dalje priprema dokument

Inicijalizacija dokumenta

- **Identifikacija** dokumenta = jednoznačno određivanje dokumenta u datom kontekstu
 - omogućava precizno referenciranje na dokument
 - stabilna i nezavisna od načina prezentacije ili fizičke lokacije
 - dokument može biti prikazan na različitim jezicima, formama (ekran, papir, ...)
 - dokument može izgledati različito za različite korisnike
 - ne mora uvek prikazivati sve informacije
 - identifikator se dodaje u metapodatke
- Primeri identifikatora:
 - interni identifikator dokumenta u okviru organizacije
 - međunarodni identifikator dokumenata (ISBN, ISSN, itd)
 - međunarodni identifikator digitalnih dela (IDDN)

Inicijalizacija dokumenta

- **Klasifikacija** dokumenta = opis karakteristika dokumenta
 - pojednostavljuje pretragu dokumenata koji se bave istim ili srodnim temama
- Različite šeme klasifikacije dokumenata
 - ISO/IEC 61355
 - ICS
 - interni šifarnici
 - ključne reči
- Metapodaci o klasifikaciji mogu da obuhvate:
 - identifikatori učesnika poslovnog procesa
 - oznake vlasnika i autora
 - funkcija dokumenta
 - jezici korišćeni u dokumentu
 - datum inicijalizacije i rok za pripremu
 - opis veza između verzija
 - prava pristupa
 - patentna i autorska prava

Priprema dokumenta

- Proizvodnja sadržaja dokumenta sve do trenutka uspostavljanja
- Počinje nakon inicijalizacije
- Metapodaci koji se dodaju u ovoj fazi bi mogli da sadrže
 - nivo razvoja dokumenta
 - ključne reči
 - rezime ili apstrakt
 - izvor dokumenta

Uspostavljanje dokumenta (establishment)

- Pre korišćenja dokument se obično **odobrava**
 - za potrebe obezbeđivanja kvaliteta
 - po pravilu se primenjuje na sve verzije
- Pravila za odobravanje se definišu na nivou
 - poslovnog procesa
 - klase dokumenta
 - pojedinačnog dokumenta

Uspostavljanje dokumenta (establishment)

- Dokument mora biti uključen u upravljanje verzijama pre odobravanja
- Metapodaci koji se dodaju u ovoj fazi
 - ID zahteva za odobrenje
 - ID podnosioca
 - datum podnošenja
 - rok za dobijanje odobrenja
 - ID osobe/organizacije zadužene za proveru
 - ID osobe/organizacije zadužene za odobravanje
 - komentari vezani za proveru i odobravanje

Korišćenje dokumenta

- Dokumenti su, sa metapodacima, dostupni za korišćenje
- Metapodaci se koriste za pretraživanje i informisanje o dokumentima i njihovim verzijama
- U metapodatke se mogu dodati komentari/iskustva korisnika o korišćenju dokumenta
- **Distribucija** = dostavljanje verzija korisnicima na kontrolisani način
 - automatsko slanje
 - obaveštavanje korisnika o dostupnom dokumentu i lokaciji
- Metapodaci vezani za distribuciju
 - distribucione liste
 - ID primalaca
 - uloge primalaca u poslovnom procesu
 - specifikacije formata distribucije
 - specifikacije formata u kojima je dokument dostupan

Revizija dokumenta

- Promena sadržaja ili promena namene dokumenta
- Obavezno u okviru upravljanja verzijama
- **Izmena sadržaja** podrazumeva novu verziju i ažuriranje metapodataka
 - prethodna verzija na kojoj se zasniva
 - verzije koje su zamenjene novom, ili se na njih utiče novom verzijom
 - ID osoba/organizacija koje su sprovele izmene
 - opis rezultata izmene
 - opis razloga uvođenja izmene
 - datum izmene

Revizija dokumenta

- Svaka verzija se objavljuje u skladu sa namenom
- Može se koristiti više verzija sve dok ispunjavaju svoju namenu
- **Povlačenje verzije** = kada se namena verzije promeni
 - izmene u metapodacima, ali ne i u sadržaju dokumenta
- Metapodaci vezani za povlačenje verzije obuhvataju
 - veze između verzija (zamenjuje/zamenjen sa)
 - verzije na koje se utiče povlačenjem
 - opis šta je učinjeno
 - opis kada je izmena načinjena

Upravljanje verzijama dokumenata

- Izmena dokumenta može da obuhvati
 - izmenu informacija u dokumentu
 - izmenu vizuelne prezentacije informacija
- Kada dokument treba da bude sačuvan kao nova verzija?
Preovlađujući stav:
 - kada se menjaju informacije - DA
 - kada se menja prezentacija - NE

Verzije i vremenska dimenzija

- Za svaku verziju postoji trenutak
 - **formiranja** - kada se verzija formira
 - **važenja** - kada se verzija smatra važećom
- Važenje verzija se može organizovati
 - sekvencijalno i
 - konkurentno

Sekvencijalno važenje verzija

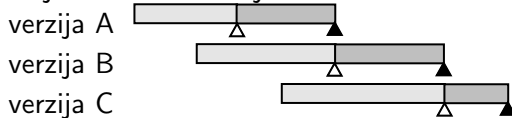
- Poslednja verzija dokumenta je jedina važeća
- Nova verzija uvek preuzima važenje od prethodne verzije
- Poslednja verzija podržava sve namene svih prethodnih verzija
- Odnos zamenjuje/zamenjen navodi se i u metapodacima odgovarajućih verzija

Konkurentno važenje verzija

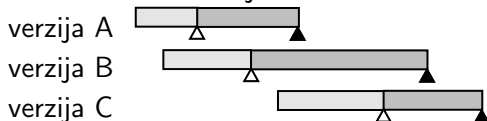
- Više različitih verzija može biti operativno u jednom trenutku
- Nova verzija ne zamenjuje automatski prethodnu u smislu važenja
- Svaka svrha pojedine verzije ostaće važeća sve do eksplicitnog ukidanja te svrhe

Sekvencijalno i konkurentno važenje verzija

sekvencijalno važeće verzije



konkurentno važeće verzije



Arhiviranje dokumenta

- Premeštanje dokumenata (verzija, metapodataka) u kompaktniju nepromenljivu formu
- Mora da ispuni ugovorne/zakonske obaveze (npr. rok čuvanja)
- Kontrolisani pristup arhivi
- Mogućnost reprodukcije dokumenata
- Nemogućnost izmena
- Arhiva = baza znanja; potrebni mehanizmi za pretraživanje
- Stabilni, nepromenljivi formati podataka

Arhiviranje dokumenta

- Metapodaci vezani za arhiviranje obuhvataju
 - prava pristupa / nivo poverljivosti
 - korišćene hardverske i softverske komponente
 - korišćeni postupci za arhiviranje i kompresiju
 - korišćeni postupci za kriptografsku zaštitu
 - digitalni potpisi
 - vremenski ciklus osvežavanja podataka (za magnetne medije)
 - istorija izmena na fizičkim nosiocima podataka
 - istorija izmena na formatu podataka
 - fizička lokacija skladišnog medija
 - fizička lokacija rezervne kopije
 - dnevnik pristupa arhiviranom dokumentu

Uklanjanje dokumenta

- Dokument se može ukloniti nakon isteka perioda za obavezno arhiviranje
- Uklanjanje sadržaja i metapodataka ne mora biti istovremeno
 - dok se drugi dokument ili verzija referišu na dati dokument trebalo bi čuvati metapodatke
- Rezultuje nepovratnim gubitkom podataka, dokumenata, relacija sa drugim dokumentima