

# Sistemi za upravljanje digitalnim dokumentima

Dragan Ivanović  
dragan.ivanovic@uns.ac.rs

Katedra za informatiku, Fakultet tehničkih nauka, Novi Sad

2015.

# Osobine standardnih poslovnih aplikacija

- Rukuju čvrsto strukturiranim podacima
- Podaci se smeštaju u relacione baze
- Korisnički interfejs prilagođen za data-intensive rad
- Podaci se razmenjuju sa spoljnim svetom preko XML formata

# Osobine standardnih poslovnih aplikacija

- Sa stanovišta korisnika: sistem rukuje dokumentima
  - primer dokumenta: faktura, izvod iz banke, nalog za plaćanje
- Sa stanovišta inženjera: sistem rukuje podacima organizovanim u tabele
- Korisnikovi „dokumenti“ su dovoljno dobro strukturirani da se mogu prikazati pomoću tabelarnih podataka
- ...bar u klasičnim domenima:  
finansijsko/materijalno poslovanje, itd.

# Osobine standardnih poslovnih aplikacija

- Sa stanovišta korisnika: dokument je i zaključak sa sednice Vlade, ili izveštaj komisije o izboru u nastavno zvanje
  - nestrukturiran, ili „slabo strukturiran“ tekst
- Sa stanovišta inženjera: dokument nije podesan za čuvanje u relacionoj bazi
- ...nova disciplina: *upravljanje digitalnim dokumentima*

# Sistem za upravljanje dokumentima

- Sistem namenjen praćenju i skladištenju digitalnih dokumenata
- Terminološka zbrka
  - document management systems
  - content management systems
  - digital asset management
  - records management systems
  - workflow systems

# Funkcije sistema za upravljanje dokumentima

- Skladištenje dokumenata
- Katalogizacija
- Pretraživanje
- Zaštita podataka
- Oporavak od katastrofe
- Arhiviranje
- Distribucija
- Upravljanje poslovnim procesima

# Skladištenje dokumenata

- Modeli kolekcije dokumenata
  - centralizovani
  - distribuirani
    - mreža diskova na različitim računarima
    - često LAN
    - jedinstveni pogled na kolekciju
    - korisnika ne opterećivati gde je zapravo dokument
- Kako obezbediti da su fajlovi dostupni i posle dugo godina
  - long-term preservation
  - sve ima svoj vek, CD, hard-disk, elektronika
  - Data formats - menjaju se standardni formati

# Katalogizacija

- U kom formatu?
- Koji su metapodaci?
- Ko obavlja katalogizaciju? Jedan čovek ili više ljudi?  
Kolaboracija



# Pretraživanje

- Velike kolekcije digitalnih dokumenata
  - <http://archive.org/>
  - <http://books.google.com/>
  - <http://arxiv.org/>
  - WWW - nije jedan sistem, ali je velika količina dokumenata i znanja
  - ...
- Forma za pretragu
- Interfejs je bitan
- ... ali nije samo on bitan
- Udaljena pretraga iz drugog sistema

# Zaštita podataka

- Autentifikacija i autorizacija
- Vrednost ovih sistema je u podacima, oni moraju biti zaštićeni
- Ako je nešto OA moraju biti rešana autorska i druga pravna pitanja

# Oporavak od katastrofe

- Backup, recovery
- Dnevni, nedeljni, mesečni
- Velika količina podataka
- Često i veliki broj korisnika, 24/7

# Arhiviranje

- Mogućnost arhiviranja nekog dokumenta
- Mora se omogućiti (često i razdvojiti) pretraga arhiviranih dokumenata
- Arhiva je baza znanja

# Distribucija

- Mehanizmi za distribuiranje digitalnih dokumenata
- Kako obavestiti javnost da je neki dokument nastao, odnosno da je promenjen?
- Ako za neki digitalni dokument niko ne zna da postoji, to je isto kao i da ne postoji

# Upravljanje poslovnim procesima

- Poslovni procesi nastajanja nekog dokumenta
- Kako se nešto arhivira?
- Ko su učesnici i ko šta radi?

# Primeri sistema za skladištenje dokumenata

- Digitalni dokumenti mogu biti skladišteni u različitim vrstama sistema koji mogu biti opšte ili usko specificirane namene
- Shodno različitim namenama, različite su i funkcionalnosti koje ovi sistemi implementiraju, kao i ciljne grupe korisnika.

# Poslovni sistemi za upravljanje sadržajima

- Sistemi za upravljanje dokumentima su sastavni deo poslovnih sistema za upravljanje sadržajem (ECMS, eng. Enterprise Content Management System)
- Pod upravljanjem sadržajem obično se podrazumeva upravljanje dokumentima, upravljanje zapisima, upravljanje veb sadržajima, upravljanje poslovnim procesima i saradnjama
- Alfresco - veb bazirani ECMS sistem otvorenog koda i otvorene arhitekture koji može da se primenjuje za upravljanje sadržajem kako u malim i srednjim organizacijama tako i u velikim geografski distribuiranim aplikacijama, obezbeđuje organizaciji sve potrebne servise za kreiranje i upravljanje digitalnim dokumentima, podržava kontrolu verzija, ima mogućnost pretrage implementirane putem Lucene biblioteke



# Digitalne biblioteke

- Nastanak digitalnih biblioteka je prirodan put prolaska ljudske civilizacije kroz informaciono doba i uvođenje savremenih tehnologija u razne aspekte ljudskog društva
- U odnosu na klasične/fizičke biblioteke, digitalne biblioteke imaju neke prednosti, a jedna od najznačajnijih prednosti je lakša diseminacija (širenje) sadržaja
- Digitalne biblioteke mogu biti opšte ili specifične namene kao što su na primer digitalne biblioteke doktorskih disertacija
- Softverske platforme otvorenog koda za kreiranje digitalnih biblioteka - DSpace i EPrints.

# Institucionalni repozitorijumi

- Softverski sistem za prikupljanje, čuvanje i preuzimanje digitalnih sadržaja koji predstavljaju digitalne dokumente relevantne za jednu instituciju
- Na primer, može se kreirati institucionalni repozitorijum naučne institucije kao što je univerzitet u kojem se skladište digitalni sadržaji kao što su radovi publikovani u časopisu, radovi prezentovani na naučnoj konferenciji, magistarske teze, doktorske disertacije, itd
- Mreže institucionalnih repozitorijuma - OpenAIRE

# Standardizacija u upravljanju dokumentima

- Standardi su raštrkani po različitim oblastima
- ISO 2709: Format za razmenu (meta)podataka
- ISO 82045: Uređenje sistema za upravljanje dokumentima
- Dublin Core: Format za reprezentaciju metapodataka
- ...

# ISO IEC 82045

- Standard za informatičko uređenje sistema za upravljanje dokumentima
- IEC 82045-1:2001 Document Management: Part 1 – Principles and Methods
  - principi i metode za definisanje metapodataka namenjenih upravljanju dokumentima u okviru životnog ciklusa
- IEC 82045-2:2004 Document Management: Part 2 – Metadata Elements and Information Reference Model
  - skup standardizovanih metapodataka za upravljanje dokumentima
  - namenjen za razmenu podataka između korisnika
  - može biti osnova za implementaciju sistema za upravljanje dokumentima
  - XML gramatika definisana DTD-om
- ISO 82045-5:2005 Document Management: Part 5 – Application of Metadata for the Construction and Facility Management Sector

# ISO IEC 82045

- Usklađenost sa standardom prema nivou podrške za rad sa verzijama
  - *Conformance class A*: podrška za sekvencijalno važeće verzije
  - *Conformance class B*: prethodno + podrška za konkurentno važeće verzije
- Usklađenost sa standardom prema stepenu pokrivanja referentnog modela podataka
  - *Conformance class 1*: Podržava samo koncept statičkog dokumenta. Nema životnog ciklusa, veza, verzija. Verzije se mogu čuvati odvojeno, bez podataka o nameni ili roku važenja.
  - *Conformance class 2*: Podrška za životni ciklus, verzije, istoriju, veze i referenciranje objekata koji se nalaze izvan sistema.
  - *Conformance class 3*: Distribucija i pretplata na dokumente u skladu sa definisanim pravilima. Podrška za arhiviranje. Referenciranje na dokumente u drugim sistemima za upravljanje dokumentima.

# Formati metapodataka

- Postoje standardizovani formati metapodataka koji se koriste u sistemima za upravljanje dokumentima
- Neki od tih standardizovanih formata metapodataka su nastali u drugim oblastima, ali su našli svoju primenu i u upravljanju digitalnim dokumentima
- Neki od formata su opšte namene (MARC 21, Dublin Core), a neki su specifične namene i mogu se koristiti samo za digitalne dokumente određenog tipa (ETD-MS)

# Dublin Core

- Standard za reprezentaciju metapodataka
- Dublin, OH, USA - mesto nastanka inicijalne verzije standarda
- Core - standard definiše osnovni skup metapodataka koja se može proširivati
- jednostavan i standardizovan skup konvencija za opisivanje objekata
- nije jedini standard u ovoj oblasti
- široko rasprostranjen

# Dublin Core

- Dublin Core ima dva nivoa
  - Simple Dublin Core: 15 elemenata za opisivanje metapodataka
  - Qualified Dublin Core: dodatna 3 elementa i mehanizam rafinacije metapodataka



# Simple Dublin Core elementi

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

# Dublin Core zapis (record)

- Skup metapodataka o nekom objektu
- Metapodaci se opisuju pomoću DC elemenata
- Svi elementi su opcionii ponovljivi

# Dublin Core zapis (record)

- DC nije vezan za konkretan format podataka
  - mogu se koristiti različiti formati
  - obično se koristi XML gramatika
- DC definiše XML namespace sa elementima koji odgovaraju DC elementima
- ove elemente možemo koristiti u okviru sopstvene XML gramatike ili nekog opšteg standarda

# Primer Dublin Core zapisa

```
<record>
  <dc:title>Na Drini ćuprija</dc:title>
  <dc:creator>Ivo Andrić</dc:creator>
  <dc:contributor>Petar Džadžić</dc:contributor>
  <dc:publisher>Srpska književna zadruga</dc:publisher>
  <dc:date>1971-01-01</dc:date>
  <dc:language>sr</dc:language>
  <dc:identifier>ISBN:0140177388</dc:identifier>
</record>
```

# Primer Dublin Core zapisa

```
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:title>UKOLN</dc:title>
<dc:description>
  UKOLN is a national focus of expertise in digital
  information management. It provides policy, research
  and awareness services to the UK library, information
  and cultural heritage communities. UKOLN is based at
  the University of Bath.
</dc:description>
<dc:publisher>UKOLN, University of Bath</dc:publisher>
<dc:identifier>http://www.ukoln.ac.uk/</dc:identifier>
</metadata>
```

# Primer DC zapisa sa pravima korišćenja

```

<record>
  <dc:title>Frog maths</dc:title>
  <dc:identifier xsi:type="dcterms:URI">http://somewhere.com/frogmaths/</>
  <dc:description>Simple maths games for 5-7 year olds.</dc:description>
  <ims:typicallearningtime>
    <ims:datetime>0000-00-00T00:15</ims:datetime>
  </ims:typicallearningtime>
  <dc:rights>
    Permission is granted for anyone to display,
    copy, modify and annotate this software.
  </dc:rights>
  <oex:rights>
    <oex:asset>
      <oex:context>
        <odd:uid idscheme="URI">http://somewhere.com/frogmaths/</odd:uid>
      </oex:context>
    </oex:asset>
    <oex:permission>
      <odd:display/>
      <odd:modify/>
      <odd:annotate/>
    </oex:permission>
  </oex:rights>
</record>

```

# Qualified Dublin Core

- Novi elementi
  - Audience
  - Provenance
  - RightsHolder
- Rafinacije elemenata
  - Alternative (npr. alternativni naslov, prevod)
  - Available (period u kome je objekat dostupan)
  - Bibliographic Citation (bibliografski opis objekta)
  - Conforms To (oznaka standarda koga objekat implementira)
  - ...

# Protokoli

- Protokoli za razmenu podataka (OAI-PMH) i protokoli za udaljeno pretraživanje kolekcije (SRU, Z39.50).
- OAI-PMH - HTTP baziran, 6 vrsta zahteva, mogućnost definisanja skupova, iterativno preuzimanje - resumptionToken
- SRU - XML, SOAP, HTTP, URI, CQL - upitni jezik
- Z39.50 - prethodnik SRU protokola, binarni protokol