



Gradivo v digitalni knjižnici

- Besedilo
- Slike
- Video
- Zvok

- Izvorno digitalni
- Digitalizirani

Digitalni informacijski viri

- Celotna besedila
 - Znanstveni časopisi
 - Odprti arhivi
 - Elektronske knjige
 - Elektronski časniki
 - Magistrska in doktorska dela
 - arhivi
- Metapodatki
 - Knjižnični katalogi
 - Bibliografske zbirke
- Drugo
 - Fotografije, slike, 3D modeli
 - Filmi, animacije
 - Zvočni posnetki

Besedilo

- Kodiranje znakov
- Strukturiranje besedila
- Omogočanje poizvedovanja

Kodiranje znakov

Različni standardi skozi zgodovino

- BCD
- EBCDIC
- ASCII
- ISO standardi (Latin-1, Latin-2...)
- Unicode (8 in 16 bitov)

Strukturiranje zapisa

- **MARC** (Machine Readable Cataloging)
- **SGML** (Standard Generalized Markup Language)
- **XML** (eXtensible Markup Language)
- **HTML** (HyperText Markup Language)
- **TEI** (Text Encoding Initiative)
-

- Omogoča boljši prikaz
- Omogoča natančnejše iskanje, npr.
 - Janez Novak
 - kot avtor
 - kot prevajalec
 - kot tema

MARC

- Standard za zapisovanje bibliografskih zapisov
- ISO 2709
- Struktura zapisa?
- 3-mestne oznake polj; podpolja
- Različni formati
- MARC21 in UNIMARC

COMARC

format izpisa

uporabniški

uporabniški polni

listek

COMARC

skrajšani

zaloga

H1

H2



1. ID=431991 M V1 22.02.2005 BFCBK::JANA
Updated: 29.02.2008 NUK::BRANKAB Copied: 12.02.2008 NUK::MILOJKAM
COBISS3: 17.04.2008 NUK::SYSTEM

001 an - nov zapis ba - tekstovno gradivo, tiskano cm - monografska
publikacija d0 - ni hierarhičnega odnosa 7ba - latinica

010 a1-55860-924-5

100 bh - publikacija z letom izida in copyrighta c2005 hslv -
slovenski ek - odrasli, zahtevno (neleposlovje) lba - latinica

1010 aeng - angleški deng - angleški

102 ausa - Združene države Amerike

105 aa - ilustracije e1 - publikacija ima kazalo

2000 aUnderstanding digital libraries fMichael Lesk

205 a2nd ed.

210 aSan Francisco cM. Kaufman Publishers dcop. 2005

215 aXXXI, 424 str. cilustr., graf. prikazi d24 cm

2251 a#The #Morgan Kaufmann series in multimedia information and systems

320 aBibliografija: str. 387-411

320 aPovzetek pri večini poglavij

320 aKazalo

6061 aDigitalne knjižnice 2NUK

6100 aelektronski informacijski viri abibliotekarstvo aknjižnice
aelektronski viri ainformacijski viri

675 c027 - Splošne knjižnice a027:025.4.036 s02 b0 vUDCMRF 2001

700 1368312419 aLesk bMichael 4070 - avtor

SGML

- Metajezik, ki omogoča specifikacijo posameznih označevalnih jezikov
- Aplikacije: XML, HTML

HTML

- Specifikacije strukture in načina prikaza
- Uporaba za spletne strani
- Tim Berners-Lee (1991)

```
<html>  
  <head>  
    <title>Pozdrav</title>  
  </head>  
  <body>  
    <p>Lepo pozdravljeni na predavanjih </p>  
  </body>  
</html>
```

XML

Format za opisovanje strukturiranih podatkov

Odprt standard, ki ga podpira W3C

```
<recipe name="bread" prep_time="5 mins" cook_time="3 hours">
  <title>Basic bread</title>
  <ingredient amount="8" unit="dL">Flour</ingredient>
  <ingredient amount="10" unit="grams">Yeast</ingredient>
  <ingredient amount="4" unit="dL" state="warm">Water</ingredient>
  <ingredient amount="1" unit="teaspoon">Salt</ingredient>
  <instructions>
    <step>Mix all ingredients together.</step>
    <step>Knead thoroughly.</step>
    <step>Cover with a cloth, and leave for one hour in warm room.</step>
    <step>Knead again.</step>
    <step>Place in a bread baking tin.</step>
    <step>Cover with a cloth, and leave for one hour in warm room.</step>
    <step>Bake in the oven at 180(degrees)C for 30 minutes.</step>
  </instructions>
</recipe>
```

TEI

Konzorcij, ki razvija in vzdržuje standard za reprezentacijo besedil v digitalni obliki predvsem za potrebe humanistike, družboslovja, lingvistike

Uporaba XML

Več:

TEI

Omogočanje poizvedovanja

- Linearno (sekvenčno iskanje)
- Invertirana datoteka
- Krmjenje
- Lingvistične metode
- Iskanje po spletu

Linearno iskanje

- Pregled datoteke od začetka do konca
- Iskanje enega ali več nizov znakov
- Prednost:
 - Velikost datoteke
- Slabost:
 - Počasnost (uporaba paralelnega procesiranja!)

Invertirana datoteka

- Princip vsebinskega kazala v knjigi
- Za vsako besedo se shrani pozicija

Prednost:

- Hitrost
- Desno krajšanje

Slabost:

- Velikost datotek

Krnjenje

- Odstanjevanje predpon in obrazil
- Rezultat je krn

- Slovnično nevtralna oblika
- Pomembno za pregibne jezike
- Zvečuje odziv

Lingvistične metode

- Analiza strukture stavkov

“I saw the man in the park with the telescope”

“Žvečilni gumi za otroke s kalcijem”

- Identifikacija fraz

- Statistične metode

Uporaba metod za avtomatsko prevajanje

Iskanje po spletu

- Žetev spletnih strani
- Razlike med iskalniki
 - Obseg
 - Rangiranje
 - Možnosti poizvedovanja

Slike

- Različni formati (GIF, TIFF, JPEG...)
- Kompresija
- Barvne, črno-bele

Video

- CD, DVD
- Izvorno digitalen/digitaliziran film
- Kompresija: MPEG

Zvok

- Glasba
- Govor
- Drugi zvoki

- Različne zahteve po kvaliteti