



ODSJEK ZA
INFORMACIJSKE ZNANOSTI
Filozofski fakultet, Sveučilište u Osijeku

DEPARTMENT OF
INFORMATION SCIENCES
Faculty of Philosophy, University of Osijek

web.ffos.hr/infoznanosti

POSTUPAK OZNAČAVANJA TEKSTA U PROJEKTIMA DIGITALIZACIJE STARE GRAĐE ZA POTREBE ZNANSTVENO-ISTRAŽIVAČKOG RADA

Boris Bosančić

Filozofski fakultet u Osijeku



20. - 21. travnja 2015.
Nacionalna i sveučilišna knjižnica u Zagrebu



PREGLED

- **Kontekst:** ZNANSTVENO-ISTRAŽIVAČKI RAD U HUMANISTIČKIM ZNANOSTIMA I PODACI ISTRAŽIVANJA
- **STANDARD ZA OZNAČAVANJE TEKSTA TEI (*Text Encoding Initiative*) KAO ALAT ZA KREIRANJE PODATAKA ISTRAŽIVANJA U HUMANISTIČKIM ZNANOSTIMA**
 - **Definirati** SPECIFIČNE POTREBE ZNANSTVENIKA KAO KORISNIKA STARE GRAĐE (KNJIGA)
 - **Razriješiti** SPECIFIČNE POTREBE ZNANSTVENIKA POMOĆU POSTUPKA OZNAČAVANJA TEKSTA
 - **Gdje?** U PROJEKTIMA DIGITALIZACIJE STARE GRAĐE...





NOVI PRISTUP ZNANSTVENO-ISTRAŽIVAČKOM RADU

- Prijelaz u 21. stoljeće OBILJEŽAVA novi pristup znanstveno-istraživačkom radu kroz povećanu uporabu interneta i informacijske i komunikacijske tehnologije.
- **e-istraživanje** (*e-Research*) u Europi i/ili **kiberinfrastruktura** (*cyberinfrastructure*) u SAD-u; **e-znanost**, **e-humanistika** i sl.
- dijeljenje i distribuiranje informacija te suradnja znanstvenika u mrežnom okruženju.
- Središnje mjesto pripada **PODACIMA ISTRAŽIVANJA!**





O PODACIMA ISTRAŽIVANJA - OPĆENITO

- Podaci na kojima se temelji znanstveno-istraživački rad:
 - **podaci znanstvenog istraživanja** (*research data*),
 - **znanstveni podaci** (*scientific data*) ili
 - samo **podaci** (*data*).
- Dijeljenje podataka znanstvenog istraživanja je temeljni element znanstvene suradnje.
- Podaci istraživanja kao središnji koncept uspostave nove informacijske infrastrukture (C. L. Borgman: *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* 2007)





DIGITALNA KNJIŽNICA PODATAKA ISTRAŽIVANJA...

(*C. Borgman, J. C. Wallis i N. Enyedy, 2006*)

- ...treba odgovoriti na sljedeća pitanja:
 - Koje su karakteristike podataka znanstvenog istraživanja? (Kojim znanstvenicima su namijenjeni?)
 - Na koji način će se moći dijeliti podaci znanstvenog istraživanja među znanstvenicima?
 - Pod kojim će se uvjetima omogućiti pristup podacima znanstvenog istraživanja?
 - Koji će oblik programske podrške rabiti digitalna knjižnica podataka istraživanja?





PODACI ISTRAŽIVANJA U HUMANISTIČKIM ZNANOSTIMA

- Predmet istraživanja u humanističkim znanostima: svaki dokument, fizički artefakt ili zapis ljudske aktivnosti.
- Izvornici podataka istraživanja: publikacije (između ostalih i stare knjige).
- Oblici/forme podataka istraživanja: tablice s numeričkim podacima o vrstama riječi zastupljenim u proučavanom tekstu, popisi osobnih imena i naziva mjesta u tekstu, locirane pogreške u tekstu i sl.
- Alati koji proizvode ili generiraju podatke istraživanja u humanističkim znanostima:
 - Rudarenje teksta (engl. *text mining*)
 - Alati za označavanje digitalnih faksimila (npr. **TILE**, **DocMark**)
 - Standardi za označavanje teksta (*Text Encoding Initiative* - **TEI**)



OPĆENITO O INICIJATIVI ZA OZNAČAVANJE TEKSTA (*TEXT ENCODING INITIATIVE – TEI*)

- Mrežno mjesto: <http://www.tei-c.org/index.xml>
- OSNUTAK: na Sveučilištu Vassar 1987. (ALLC, ACH i ACL) - jedan od najznačajnijih događaja u razvoju digitalne humanistike (S. Hockey, W. McCarthy i dr.)
- CILJEVI: osigurati uspješnu razmjenu humanističkih tekstova u znanstveno-istraživačkom radu & standardizirati principe označavanja teksta
- Razvoj TEI smjernica/vodiča za označavanje teksta – od P1 (1990) do P5 (2007)
- TEI PROJEKTI: <http://www.tei-c.org/Activities/Projects/>





TEI PROJEKTI U SVIJETU

Perseus Digital Library

(Sveučilište Harvard, 1987)



Orlando projekt

(Sveučilišta Alberta i Guelph, 1996)



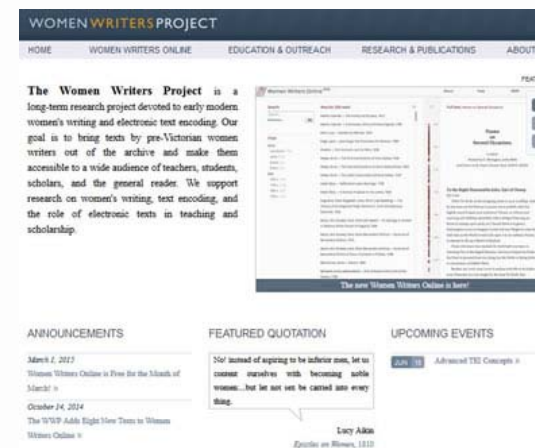
Oxford Text Archive (OTA)

(Sveučilišta Oxford, 1976)



The Women Writers Project (WWP)

(Sveučilište Brown, Northeastern 1988)





TEI PROJEKTI U RH

Hrvatski jezični korpus

(Institut za hrvatski jezik i jezikoslovlje, 2011)



Croatiae auctores Latini

(Filozofski fakultet u Zagrebu, 2014)



Croatiae auctores Latini (CroALa)
Collectio electronica

Qui et quid? Opera Auctores Schola et opuscula Nexus et bibliographica Textus libri

Caroline editio
Tome I-IV (Zagreb)
Tome V-VIII (Zagreb, 1881, 1882, 1883, 1884)
Tome IX-XIII (Zagreb, 1885, 1886, 1887, 1888, 1889, 1890)

1881-1890

Facultet Philosophiae Universitatis Zagrebensis
Zagreb, n. 4, 11. Rue Fenech et D. 48202
ISBN: 978-953-373-256-2

ERGO INARE QUAEQUE

IN COLLECTIOE HUIUS LIBRI: EDITIONES, ACCURATAE, AUCTORES (PROFANAE) ET
DE OMNI FINE ET UTI DE OMNI (1881)

QUAE ET LAETITIAE, LIBERUM DIGNITATEM

EDICIJA: digitalna knjižnica hrvatske tiskane baštine

(Filozofskifakultet u Osijeku, 2010)



20. - 21. travnja 2015.
Nacionalna i sveučilišna knjižnica u Zagrebu



ZNANSTVENE ZAJEDNICE ZAINTERESIRANE ZA PROUČAVANJE STARE GRAĐE

- Povjesničari knjige
- Povjesničari umjetnosti
- Povjesničari književnosti
- Sociolozi
- Lingvisti
- Filolozi
- Bibliografi
- Stručnjaci zaštite građe
- Knjižničari (pogotovo u specijalnim knjižnicama)





ZNANSTVENICI KAO KORISNICI STARE GRAĐE

- Stare knjige se proučavaju u kontekstu sljedećih aspekata:
 - **društvenog** (duhovna klima vremena, kulturne i druge posljedice, profil čitatelja i sl.);
 - **knjižarskog** (prodaja, cijena, rasprostranjenost i sl.);
 - **književno-sadržajnog** (književne interpretacije, usporedna čitanja, sekundarni tekstovi pored primarnog poput posvete, komentara i sl.);
 - **lingvističkog** (vrste riječi, grafijski i fonemski sustavi i sl.);
 - **nakladničko-grafičarskog** (tehnike tiskanja, izrada pomičnih slova i sl.);
 - **materijalnog** (obraćanje pažnje na samu knjigu i sastavne elemente poput uveza, materijala od kojeg je sačinjena, formata i sl.).





PRIMJER: povjesničari knjige...

- ŠTO PROUČAVAJU? povijesne likove i događaje, duhovnu klimu vremena, kulturne i druge posljedice, profil čitatelja, profil autora itd.
- S označiteljskog stajališta povjesničari knjige mogu biti zainteresirani za sljedeća svojstva ili obilježja teksta:
 - postojanje različitih zasebnih elemenata sadržaja (npr. tiskarski i vodeni znakovi, ex librisi i sl.)
 - bilješke autora i čitatelja,
 - greške u tekstu uključujući i nečitljiv tekst,
 - osobna imena, nazivi mjesta, zemljopisni nazivi i dr.
 - Primjerice, povjesničar knjige *Z. Velagić* proučavajući djelovanje nekoliko hrvatskih autora 18. stoljeća na hrvatskom sjeveru ispituje njihov odnos prema pisanju, knjizi i čitatelju. (Velagić, 1999)





SPECIFIČNE POTREBE ZNANSTVENIKA

- Razlikuju se dvije vrste:
 - bibliografskog karaktera (informacijske potrebe uglavnom vezane uz knjižnični katalog)
 - nebibliografskog karaktera (vezane uz tekst)
- Razriješene specifične potrebe znanstvenika javljaju se kao **podaci istraživanja** u novoj **znanstvenoj informacijskoj infrastrukturi**.
- Standard za označavanje teksta (**TEI**) prepoznaje se kao alat za razrješavanje specifičnih potreba znanstvenika.





SPECIFIČNE POTREBE ZNANSTVENIKA NEBIBLIOGRAFSKOG KARAKTERA KAO KORISNIKA STARIH KNJIGA

- **potrebe** vezane uz obavljanje specifičnih postupaka kako bi se proizveli **podaci znanstvenog istraživanja**
- PRIMJERI:
 - izraditi popis stranih riječi koje se javljaju u tekstu
 - označiti osobna imena i mjesta u tekstu
 - utvrditi tragove interakcije čitatelja s tekstom (bilješke, podcrtanost, korištenje interpunkcije)
 - označiti vrste riječi (imenice, glagole i sl.)
 - označiti i prebrojati grafeme u tekstu
 - označiti i prebrojati ligature u glagoljičkom tekstu i dr.





PRECIZIRANJE SPECIFIČNIH POTREBA ZNANSTVENIKA ZA POTREBE POSTUPKA OZNAČAVANJA TEKSTA

- Specifični postupci kojima se dobivaju/proizvode podaci znanstvenog istraživanja: *prebrojavanje, označavanje, uspoređivanje, izrada, razvrstavanje, sortiranje, pretraživanje, pregledavanje* i dr.
- Opis specifične potrebe znanstvenika (SPZ)
- Način dosadašnjeg razrješavanja SPZ
- Elementi SPZ koji trebaju biti označeni
- Praktična svrha SPZ
- Precizirana SPZ u obliku u kojem se može razriješiti postupkom označavanja teksta pomoću TEI-a





PRIMJERI...

RED. BROJ	OPIS SPECIFIČNE POTREBE (SPZ)	NAČIN DOSADAŠNJEG RAZRJEŠAVANJA SPZ	ELEMENTI SPZ KOJI TREBAJU BITI OZNAČENI	PRAKTIČNA SVRHA SPZ	PRECIZIRANA SPZ ZA POTREBE POSTUPKA OZNAČAVANJA TEKSTA POMOĆU TEI-A
1.	„Označiti osobna imena i mjesta te ih povezati u odgovarajući kontrolirani rječnik termina s dodatnim informacijama“. „Mogućnost prepoznavanja imena, mjesta ili nekih dijelova rukopisnih zabilješki“ u tekstu.	Pregledavanjem knjiga.	osobna imena, mjesta, bilješke*, ex librisi**	PREGLEDAVANJE	PREGLEDAVATI prema osobnim imenima, mjestima, bilješkama i ex librisima.
2.	Označiti i prebrojati grafeme u tekstu. Opisati i razvrstati označene grafeme prema unaprijed određenim obilježjima ili bez njih.	„Ručno“ prebrojavanje i razvrstavanje (grafema, vrsta riječi i sl.).	grafemi, obilježja grafema	PREBROJAVANJE PREGLEDAVANJE	PREBROJATI/ PREGLEDAVATI označene grafeme u tekstu prema unaprijed određenim ili neodređenim obilježjima.



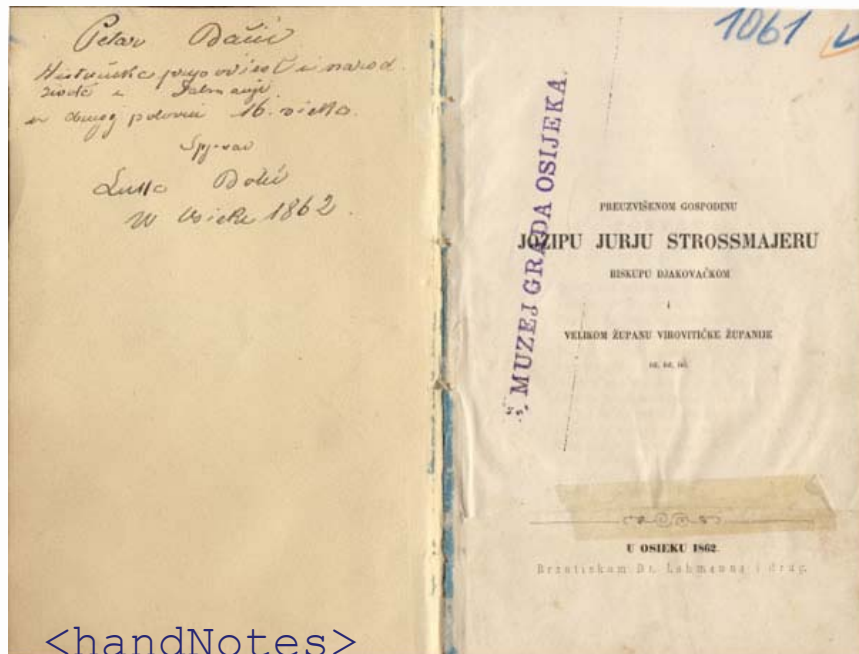
PRILAGODBA TEI STANDARDA KNJIŽNIČNIM POTREBAMA (U PROJEKTIMA DIGITALIZACIJE GRAĐE)

- *Best Practices for TEI in Libraries* -
<http://www.tei-c.org/SIG/Libraries/teiinlibraries/main-driver.html>
- **TEI knjižnična interesna skupina**
- RAZINE OZNAČAVANJA TEKSTA:
 - **Razina 1 (Level 1)** – tekst podređen svom digitalnom faksimilu, automatska konverzija i označavanje teksta.
 - **Razina 2 (Level 2)** – i dalje tekst podređen svom digitalnom faksimilu; označavanja naslova i podjela teksta na sastavne dijelove; minimalno označavanje teksta.
 - **Razina 3 (Level 3)** – jednostavna analiza; uključuje mnoge elemente iz *osnovnog* modula TEI vodiča.
 - **Razina 4 (Level 4)** - ispravljanje teksta; tekst može stajati samostalno bez pripadnog digitalnog faksimila;
 - **Razina 5 (Level 5)** – u postupku označavanja teksta sudjeluju predmetni stručnjaci područja kojem tekst pripada.





NASLOVNE STRANICE S POSVETOM, RUKOM PISANOM BILJEŠKOM, PEČATOM I SL.

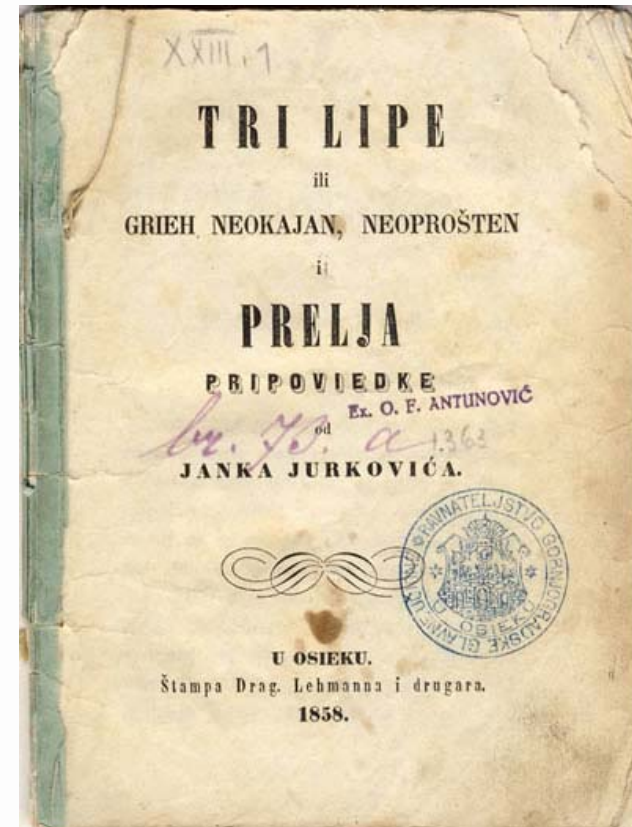


<handNotes>

<handNote xml:id="pb" script="handwritten" medium="pencil">

<p>Bilješka nepoznatog autora</p>

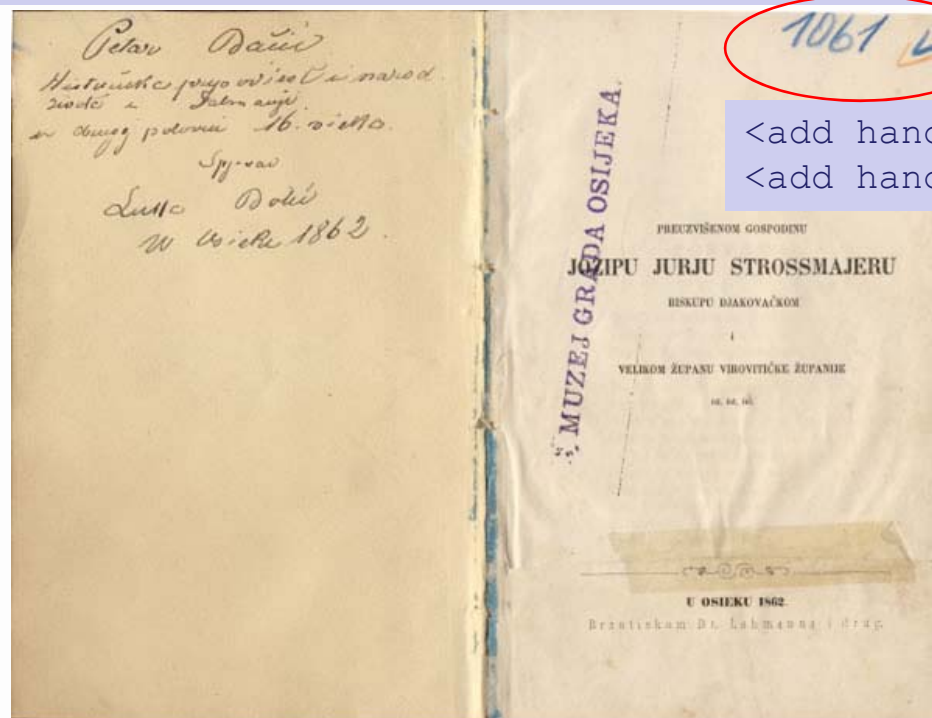
</handNote>





... Drugi primjer ... nastavak ...

```
<add place="top" hand="#lib-sig-1_or_inv-num-1">  
<seg exclude="#inv-num-1" type="old_library_signature" xml:id="lib-sig-1">1061</seg>  
</add><seg copyOf="#lib-sig-1" exclude="#lib-sig-1"  
type="inventory_number" xml:id="inv_num-1"/>
```



```
<add hand="#rev1">✓</add>  
<add hand="#rev2">✓</add>
```

```
<handNote xml:id="lib-sig-1_or_inv-num-1" scribe="Auditor1"  
script="handwritten" medium="blue_felt_pen">  
<p>Stara knjižnična signatura ili inventarni broj upisan od  
pretpostavljenog Revisora 1</p></handNote>
```



DIGITALNI FAKSIMIL S ILUSTRACIJOM U TEKSTU/SADRŽAJU.

POVEZATI ilustracije u tekstu s elementima teksta koji se na nju odnose

```
<zone
xml:id="pict-eva"
ulx="233"
uly="225"
lrx="297"
lry="370" >
<desc>Eva</desc>
</zone>
```



```
<persName ref="#eva">Eva</persName>
```

```
<link targets="#eva #pict-eva"/>
```



OCJENA KORISNOSTI I ISPLATIVOSTI OZNAČAVANJA TEKSTA STARIH KNJIGA NA HRVATSKOM JEZIKU (doktorsko istraživanje iz 2011)

Redni broj ispitanika	Ocjena korisnosti	Ocjena isplativosti	Jeste li voljni koristiti TEI u svom znanstvenom radu?
1	5	2	DA, UZ STRUČNJAKA ZA OZNAČAVANJE TEKSTA
2.	5	3	DA, UZ STRUČNJAKA ZA OZNAČAVANJE TEKSTA
3.	5	5	DA, SAMOSTALNO
4.	5	3	DA, UZ STRUČNJAKA ZA OZNAČAVANJE TEKSTA
5.	3	1	NISAM SIGURAN/SIGURNA (MOŽDA U BUDUĆNOSTI)
6.	5	4	DA, UZ STRUČNJAKA ZA OZNAČAVANJE TEKSTA
7.	5	3	DA, SAMOSTALNO
	4,71	3,00	





TEORIJSKI OKVIR ZA PROVEDBU POSTUPKA OZNAČAVANJA TEKSTA





ZAKLJUČAK

- ZNANSTVENO-ISTRAŽIVAČKI RAD u digitalnom okruženju dovodi do nastajanja nove znanstvene informacijske infrastrukture (DIGITALNE KNJIŽNICE predstavljaju njene dijelove!) koja se temelji na PODACIMA ISTRAŽIVANJA.
- SPECIFIČNE POTREBE ZNANSTVENIKA u proučavanju stare građe tiču se dobivanja različitih PODATAKA ISTRAŽIVANJA
- PROJEKTI DIGITALIZACIJE STARE GRAĐE – implementacija postupka označavanja teksta pomoću TEI standarda kao jedan od načina dobivanja PODATAKA ISTRAŽIVANJA!
- Postupak označavanja teksta pomoću TEI-a koristan je, ali ne i u istoj mjeri isplativ postupak razrješavanja specifičnih potreba znanstvenika!
- Potreba **veće suradnje** humanističkih i informacijskih stručnjaka!





HVALA NA PAŽNJI!
bbosancic@ffos.hr

