

OZNAČIVANJE s DocMark¹-om :

- početni impuls u istraživanjima stare i rijetke građe

Prof.dr.sc. Mario Essert, Doc.dr.sc. Marijana Tomić
messert@fsb.hr, mtomic@unizd.hr

Fakultet strojarstva i brodogradnje, Zagreb
Odjel za informacijske znanosti, Zadar

20. travnja 2015.

Sadržaj:

1 Matematička funkcija (preslikavanje)

- bez nje se ne može ☺
- + internet koji sve povezuje

2 TEIMark

- $Y = \text{tekst}$
- $X = \text{HKB}, Y = g, t, o, \dots$

3 Hrvatsko jezikoslovlje

- makro funkcije (korpus, atlas)
- mikro funkcije (morfologija)

4 Hrvatska leksikografija

- tvorba rječnika
- sveza leksema (LLOD)

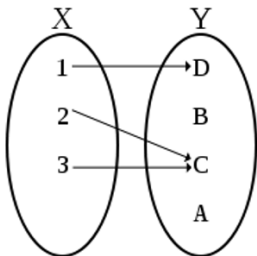
U temelju DocMark vizualizacije je *matematička funkcija!*

domena = {1,2,3}

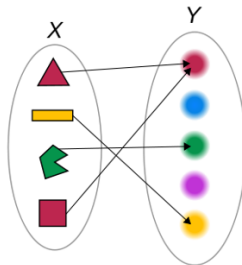
kodomena = {A,B,C,D}

domena={oblik}

kodomena={boja}



$f1 = \{(1,D), (2,C), (3,C)\}$



$f2 = \{(trokut, ljubičasta), (pravokutnik, žuta), \dots\}$

Marijana birala **X** i **Y**, a '**dečki**' (Nikola, Vlado, Mario) 'štrikali' funkcije

- **DocMark domena** - paleografska, kodikološka, tipografska, jezična, ... obilježja (elementi, podskupovi)
- **DocMark kodomena** - vizualne oznake (oblika, uzoraka, boja)
- A sve ostalo - **kompozicijom** (*nadovezom*) funkcijâ

Shema \Rightarrow OZNAKA (tag) = kompozicija 'broja' i boje

Arak						
	Arapski broj		S7_11	9	010	orange
	Rimski broj		S7_12	🔥	041	orange
	Glagoljički broj		S7_13	🔌	165	orange
	Latinica		S7_14	A	179	orange
Folija unutar arka						
	Arapski broj		S7_21	9	010	blue
	Rimski broj		S7_22	🔥	041	blue
	Glagoljički broj		S7_23	🔌	165	blue
	Latinica		S7_24	A	179	blue
Folijacija						
	Arapski broj		S7_31	9	010	green
	Rimski broj		S7_32	🔥	041	green
	Glagoljički broj		S7_33	🔌	165	green
	Latinica		S7_34	A	179	green
Paginacija						
	Arapski broj		S7_41	9	010	red
	Rimski broj		S7_42	🔥	041	red

Zadano: **X1: Broj**, **X2: Geometrija**; Traži se: $f_1(X_1) \cdot f_2(X_2)$; Na primjer: 'glagoljički broj' -od- (folije unutar arka) = **plava** sklopka (dipinto di blu 😊)

- **X1** - Arapski, Rimski, Glagoljički, Latinica
- **Y1** - 9, plamen, sklopka, A
- **X2** - Arak, Folija unutar arka, Folijacija, Paginacija
- **Y2** - narančasto, plavo, zeleno, crveno

Otvorite: <http://docmark.cingel.hr>

Parent: Inicijal > Latinka > 1 redak
Active marker: Inicijal > s ukrasnom crvenj | [x]

Prijava

Login

test1

Password

Remember me

Log in

DocMark

Knjige

Naziv	Broj Stranica	Korisnik
Casoslov opatica Cika	2	Marijana Tomić
Mazec brevarij_baja	5	Marijana Tomić
Dvusi novljanski brevarij (SSSRH)	2	Marijana Tomić
HDA_cimser_1	2	Nikola Glumac
HDA_cimser_2	2	Nikola Glumac
HDA_cimser_3	3	Nikola Glumac
nasional	2	Test User
CroatSola	1	Mario Lončarić
Ikonuabula_oroba	1	Test User
Vjebta INFO ZNANOSTI	0	Test User

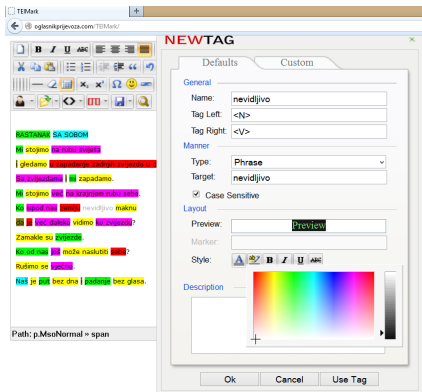
Uvoste za pokretanje

Alatna traka

20 korisničkih računa: test1, test2, ... test20 /zaporka: 123456

- DocMark - probajte sami (na par knjigica/slikica), i razmišljajte o primjeni, nisam Vam potreban 😊!

- isto, samo se kodomena **Y** promijenila 😊!

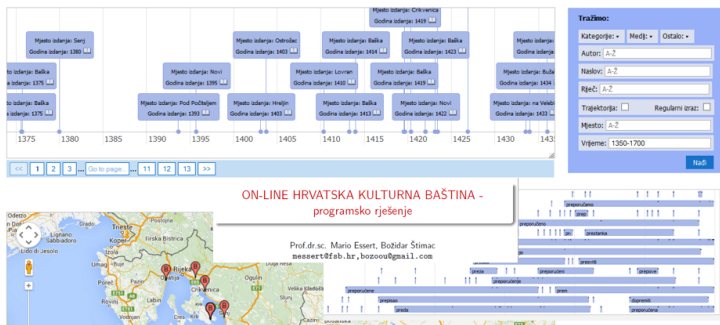


Vizualni alat za *Borisove* mučne poslove

- umjesto digitalizirane slike imamo digitalizirana slova (tekst)!
- definiranje vlastitih oznaka (u hijerarhijskim skupinama)
- povezivanje (*kompozicija funkcija*) do bilo koje dubine
- ručni i poluautomatski rad
- uvoz/izvoz u poznatim ili posebnim XML formatima
- lako i brzo pretraživanje/zamjena
- Pogledajmo (kratko!) filmić

TEI - text encoding initiative
<http://www.tei-c.org>

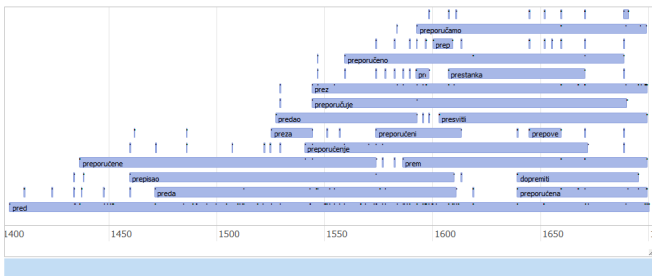
X = {mediji (tekst,slika,zvuk)}, **Y** = {prostor, vrijeme}



Y = mrežni okvir **Hrvatske kulturne baštine** (realizacija: *Božidar Štimac*)

- **Vremenski** - unutar vremenske osi od najstarijeg do modernog doba,
- **Prostorno** - na *google*-ovim kartama, preko povijesnih areala
- **Obilježjima** - oznakama, specifičnim za područje istraživanja
- **Leksikografski** - svezom pojmova preko on-line rječnika/enciklopedija
- **Tražilicom** - povezivanje (kompozicija) po višestrukim kriterijima.

Makro funkcije: od dokumenata do riječi u vremenu



Tražimo:

Kategorije: **•** Medj: **•** Ostalo: **•**

Autor:

Naslov:

Rječ:

Trajektorija: Regularni izraz:

Mjesto:

Vrijeme:

bi božja volja bila, da bi moje hčere **preminule** i njihju porod ale bližnja moja parenta,

od ovoga svita, i pristavi biše prizvani **pred** podknežina s Olovine, ki side na stoli

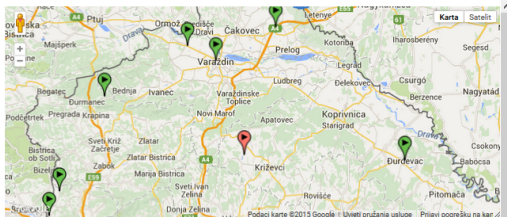
g(ospo)d(())na plovana Jurja novogradskega vikara i arhipreda vinodolskega, ka bista velika pomočnika

Praćenje hrvatskih riječi u vremenu ($X = \{\text{riječi}\}$, $Y = \{\text{vrijeme, knjiga}\}$)

- **Traženje niza znakova** - i njegova prikazba na vremenskoj osi
- **Ispis rečenica** - u kojima se izraz nalazi, u knjizi/listini
- projekt **Arealno-dijakronijski korpus**, voditelj *dr.sc. Jurica Budja*, IHJJ, ZG

ili : od hrvatskih (iz)govora do prostornih oznaka

Zvučni atlas hrvatskih govora

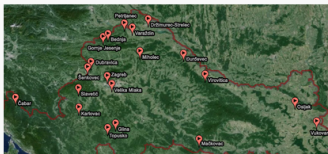


psl.jat
psl.*zdj,*skj,*stj
psl.d i dej
h
dočetno l
psl.e
psl.q
akcentuacija
...

1. Jesi li danas kupila mlijeko?
2. Tjesto mora biti mekano.
3. Na ovom mjestu je nekada bila štala.
4. Nemam baš vremena, ali možemo proboriti koju riječ.
5. Moj djed je uvijek govorio tako.
6. Uzmi to od djeteta.

Opće | Razglednica

Ovdje može ići proizvoljni opis...te se dodavati proizvoljne slike...



Praćenje hrvatskih riječi/izraza u prostoru (**X** = {riječi}, **Y** = {prostor})

- **Zvučni atlas hrvatskih govora** - voditelj projekta: *prof.dr.sc. Velimir Piškorec*, Filozofski fakultet, Zagreb
- Poslušajmo malo hrvatskoga **dijalektološkog blaga!**

Mikro funkcije: od slogova i glasovnih promjena do riječi

Morfologija

Imenice Pridjevi Glagoli Baza Slogovi

Unesite imenicu te odaberite rod i deklinaciju.

Imenica

Rod

Deklinacija

Napredno

Zivo

Glasovne promjene

Produljena osnova

realizacija: *Joško Markučić*

Jednina

Padež	<input type="checkbox"/> Odaberi	<input checked="" type="checkbox"/> Odaberi	F
N	festival	festival	E
G	festiva	festivala	E
D	festivu	festivalu	E
A	festiv, festiva	festival, festivala	E
V	festive, festivu, festiv	festivale, festivalu, festival	E
L	festivu	festivalu	E
I	festivom, festivem	festivalom, festivalem	E

Mnozina

Padež	<input type="checkbox"/> Odaberi	<input type="checkbox"/> Odaberi	<input type="checkbox"/> Odaberi	<input checked="" type="checkbox"/> Odaberi	<input type="checkbox"/> Odaberi	<input type="checkbox"/> Odaberi
N	festivi	festivovi	festivevi	festivali	festivalovi	festivevi
G	festivala	festivova	festiveva	festivala	festivalova	festiveva
D	festivima	festivovima	festivevima	festivalima	festivalovima	festivevima
A	festive	festivove	festiveve	festivale	festivalove	festiveve
V	festivi	festivovi	festivevi	festivali	festivalovi	festivevi
L	festivima	festivovima	festivevima	festivalima	festivalovima	festivevima
I	festivima	festivovima	festivevima	festivalima	festivalovima	festivevima

Hrvatsko oblikoslovlje : tvorba imenica, pridjeva i glagola; i to on-line, jupiii!

- Upis oblikovane riječi u bazu + moguća promjena/brisanje
- Baza/leksikon s oko milijun riječi (obitelj *Pinjatelya*, Zadar 2001.)

Relacijski: od dokumenta (datoteke) - do rječnika

Datoteke:

- Skup A
 - HAZU
 - Kranjčević
 - Bugarkinje.txt
 - Ela_o_Silvju.txt
 - Izabrane_pjesme.txt
 - Pjesme.txt
 - Matoš
 - Matos_Cvijet_sa_Ra
 - TestCache
 - Test
 - croatianStopWords.txt
- Operacija
 - A
 - B
 - A U B
 - A ∩ B
 - A-B
 - B-A
 - A*B
 - Skup B
 - HAZU
 - Kranjčević
 - Bugarkinje.txt

Rječnik Konkordancija Usporedba Uređivanje Zamjena/proširivanje Postavke

Prikaži rječnik Spremi rječnik Radnje

Rastavi na slogove
 Sortiraj stupac "A..z" obrtimično

Označi	Rbr	SVE A..z	a b c č ć d đ e f g h i j k l m n o p r	Pojavak	Oznaka/e
<input checked="" type="checkbox"/>	1	baba lema: ba	1	izjn0, izmg0, immg0	
<input checked="" type="checkbox"/>	2	hljeba lema: hljeb	3	imjg0	
<input checked="" type="checkbox"/>	3	neba lema: nebo	10	isjg0, ismno, ismg0, isma0, ismv0	
<input checked="" type="checkbox"/>	4	treba lema: trebati	9	gp3j	
<input checked="" type="checkbox"/>	5	bomba lema: bomba	2	izjn0, izmg0	
<input checked="" type="checkbox"/>	6	doba lema: doba	4	isjn0, ismg0	
<input checked="" type="checkbox"/>	7	zloba lema: zloba	7	izjn0, izmg0	
<input checked="" type="checkbox"/>	8	roba lema: rob	8	imjg0, imja0, izjn0, izmg0	

Automatska on-line tvorba abecedarija (i uskoro, rječnika/leksikona)

- čestotnost riječi, abecedno i 'gramatičko' slaganje
- unaprijedno i obrtimično; rastuće i padajuće
- usporedbe dokumenata, logičke operacije nad rječnicima, obradbe, napredne konkordancije, *PDF rječnik*, (realizacija: *Dražen Mrvoš, Jakov Topić*)

Sveze s leksikografskim blagom (LZMK, IHJJ, HAZU)

TEIDIC **CTIONARY**

- D struna od ranije
- D linkovi
- D filmovi
- D baza_struna
- D hazu-out

list

CTIONARY: linkovi (vidi više)

list

CTIONARY: baza_struna (vidi više)

formatirani dio filma, folije ili plo?e

Import new dictionary

TEIDIC **stijenska**

- D struna od ranije
- D linkovi
- D filmovi
- D baza_struna
- D hazu-out

CTIONARY: struna od ranije (vidi više)

unutarnja **površina** ili **pregrada** kakva prostora

površina (close all) x

CTIONARY: struna od ranije (vidi više)

ostvarena **ploha** koja zatvara **bjelo** prema okolnome prostoru

pregrada (close all) x

CTIONARY: struna od ranije (vidi više)

okomita **stijena** koja unutarnji dio broda dijeli **u** pregratke

$X = \{\text{riječi}\}$, $Y = \{\text{on-line HR-rječnici}\}$

- Dohvaćanje pojmova iz on-line rječnika za riječi u dokumentu,
- rekurzivna povezanost do bilo koje dubine
- izvlačenje semantičke informacije (realizacija: *Juraj Benić*)
- dodavanje vlastitih on-line rječnika/enciklopedija,
- gradi se **CroLLOD** triplestore - hrvatski jezikoslovni povezani podaci (+ ugradnja mrežnih ontologija, **Ivana Sarić-Dević**, Split)

lako bez projekata, nagrada, novaca ... ↙ MI *letimo* 😊 ↘



NSK 5. festival: *Nel blu dipinto di blu* - VOLARE

- Volare, o-o... Cantare, o-oo-o...
- *Nel blu dipinto di blu* → Felice di stare lassù.
- *Nel blu degli occhi tuoi blu*
- Felice di stare quaggiù → *Sretan da sam tu ...!*
- **Hvala!**