

arnes   
povezujemo znanje



**MREŽA ZNANJA**

Ljubljana, 3.–5. december 2024

---

# Raziskovalni program Digitalna humanistika: viri, orodja in metode

Filip Dobranić

Inštitut za novejšo zgodovino



# Digitalna humanistika?

Digitalna humanistika (DH) je področje raziskovanja na preseku [...] digitalnih tehnologij in humanističnimi disciplinami. Obsega tako sistematično rabo digitalnih resursov v humanistiki kot analizo njihove rabe.

Wikipedia, CC Attribution-ShareAlike 4.0

- Od rojstva transdisciplinarna.
- Digitalna v več pogledih:
  - ▶ digital(izira)ni viri,
  - ▶ digital(izira)ne metode.
- V Sloveniji zgolj formalno nekaj novega.

---

# Zakaj DH potrebuje SLING?

Digitalna humanistika potrebuje svojo infrastrukturo. Knjižnice, arhive, orodjarne zemljevide nevarnih področij, teste in réperje.





## Kaj delamo?

- Ustvarjamo *kompleksne digitalne edicije*
  - ▶ Procesiramo kompleksne dokumente
  - ▶ Procesiramo zvočne zapise oralne zgodovine
  - ▶ Procesiramo slike in fotografije
- Razvijamo metode za slovenske DH zbirke
  - ▶ NLP metode za pisne in ustne zbirke
  - ▶ Metode za procesiranje večjezičnih zbirk
  - ▶ Metode za bogatenje besedil
- Razvijamo mešane za analizo DH podatkov
  - ▶ Mešane metode za analizo omrežij
  - ▶ Mešane metode za prostorske analize
  - ▶ Mešane metode za časovne analize

## Kdo delamo?

- Inštitut za novejšo zgodovino
- Fakulteta za računalništvo in informatiko UL
  - ▶ Laboratorij za računalniško grafiko in multimedije
- Filozofska fakulteta UL
  - ▶ Oddelek za prevajalstvo

# Pričakovanja



## Periode

- 18e eeuw (20)
- 19e eeuw (1219)
- 20e eeuw (89770)
- 21e eeuw (2124)

Kies periode...

## Soort bericht

- Advertentie (2561)
- Artikel (83286)
- Familiebericht (97)
- Illustratie met onderschrift (7189)

## Verspreidingsgebied

- Landelijk (51924)
- Nederlands-Indië / Indonesië (1980)
- Nederlandse Antillen (2736)
- Regionaal/lokaal (36218)
- Suriname (236)
- Verenigde Staten (39)



## Koningin Beatrix' eerste jubileum

Beatrix de vorm van een symbolische cheque aan de voorzitters van het Koningin Juliana Fonds en het Koningin Beatrix Fonds cadeau hebben gedaan. •Staatssecretaris Neelie Smit-Kroes

Krantentitel

De Telegraaf

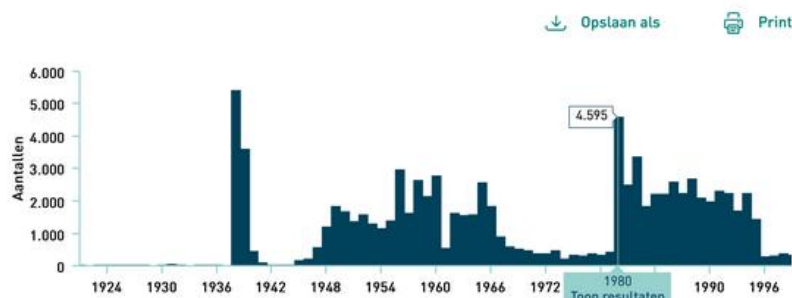
Datum

30-04-1981

Meer details



## Grafische weergave van zoekresultaten 89.445 krantenartikelen



Onderstaande tekst is niet 100% betrouwbaar

Advertentie

Advertentie

Koningin Beatrix jarig

★ Koningin Beatrix viert vandaag in huiselijke kring haar 51-ste verjaardag. Op Koninginnedag, 30 april, zal zij een bezoek brengen aan de Zuidhollandse gemeenten OudBeijerland en Goedereede. De NOS-televisie herhaalt ter ere van de verjaardag van de vorstin vanavond de documentaire 'Beatrix, koningin'. Dit televisie-portret werd vorig jaar gemaakt ter ere van haar vijftigste verjaardag en oogste veel waardering. 'Beatrix, koningin' is te zien op Nederland 3 en begint na het journaal van acht uur.

Beleggers willen Van Goghs geboortehuis kopen

Meester werken weer veilig

Gorbatsjov naar Bonn

Toon alle artikelen | Verberg alle artikelen





# Resničnost digitalizacije zgodovinske periodike

## Kaj želimo

- Ploskovne posnetke (skene)
- OCR (razpoznavo besedila/znakov)
- Metapodatke o postavitvi vsebine (layout)
- Označene nebesedilne vsebine

## Ali to imamo?

- Ja.
- Ja, ampak.
- Ne.
- Ne.



---

## Prvi korak: sPeriodika

### Kako se je začelo?

- 250.000 izvodov od ~1800 do 1914.
- Digitalizirani v različnih časovnih obdobjih.
- V različnih jezikih.
- S kar nekaj "zatipki".

### Kaj smo morali narediti?

- Filtrirati glede na jezik.
- Ločiti kvalitetno "zrno" od "plevela".
- Filtrirati glede na jezik.
- Nežno popraviti napake OCR.



CONCORDANCE

Slovenska periodika (1771-1914)

simple Lendava ● 152  
0.17 per million tokens ● 0.000017%

🔍 ⬇️ ☰ 👁 📄 ✂ ⚙ ⚙ ⚙ ⚙ ⚙ ⚙ ⚙ KWIC ⓘ

Details    Left context    KWIC    Right context

1	<input type="checkbox"/>	📘	Zvonček (Lju... : Vidu Pikuševa iz Dolnje <b>Lendave</b> ; Ivan, Marjan in Matko Sv	📄
2	<input type="checkbox"/>	📘	Družba sv. Moho... Mih., Šupi Ivan.4 Gornja <b>Lendava</b> . Recek Jurij, Štercin Blaž,	📄
3	<input type="checkbox"/>	📘	Društvo Pravniki... v» -š e k Srečko (Dolnja <b>Lendava</b> ), A r k o Josip (Kočevje),	📄
4	<input type="checkbox"/>	📘	Katoliško tisko... ar Jakob, župnik v Gornji <b>Lendavi</b> in zatem v Črensovcih, ur	📄
5	<input type="checkbox"/>	📘	Katoliško tisko... ef, tedaj kapelan v Dolnji <b>Lendavi</b> , spisoval je obširne »Zgo	📄
6	<input type="checkbox"/>	📘	Klekl Jožef ● N... de Kermedina ali pa od <b>Lendave</b> do Egersega. Neščem to	📄
7	<input type="checkbox"/>	📘	Klekl Jožef ● N... kakša je od Radgone do <b>Lendave</b> . Po tej poti smo mimo 5	📄
8	<input type="checkbox"/>	📘	Katoliško tisko... k Madjare, ki so pridrli iz <b>Lendave</b> . In res so 1. 1704. pridrli	📄
9	<input type="checkbox"/>	📘	Matica Slovensk... eri štirsko krajevno imš: <b>Lendava</b> , f. „144. „ Mäk je furlansl	📄
10	<input type="checkbox"/>	📘	Ljudska tiskarn... šzecs, graščaka v Gornji <b>Lendavi</b> , Jarija pl. Weissenecka, c	📄
11	<input type="checkbox"/>	📘	Ljudska tiskarn... eto, roj. Szecs iz Gornje <b>Lendave</b> . 2.Kedar bode kateri sin	📄
12	<input type="checkbox"/>	📘	Ljudska tiskarn... o uro od Tišine), v Dolnji <b>Lendavi</b> (?e madjarska, pa še mnc	📄
13	<input type="checkbox"/>	📘	Ljudska tiskarn... :transka proga do Dolnje <b>Lendave</b> , koder naj izstopi, in če b	📄
14	<input type="checkbox"/>	📘	Matica Slovensk... ak Jožef, kaplan v Gornji <b>Lendavi</b> . Borovnjak Jožef, župnik '	📄
15	<input type="checkbox"/>	📘	Matica Slovensk... tar Jožef, župnik v Gornji <b>Lendavi</b> . Skrlec Jožef, gostilničar	📄
16	<input type="checkbox"/>	📘	Matica Slovensk... ti F el so Lendva, Gornja <b>Lendava</b> (OberLimbach). Na eni st	📄
17	<input type="checkbox"/>	📘	Matica Slovensk... iti Sveti Jurij blizo gornje <b>Lendave</b> , ali kaj je Csop, ki še na r	📄
18	<input type="checkbox"/>	📘	Matica Slovensk... ak Jožef, kaplan v Gornji <b>Lendavi</b> . Borovnjak Jožef, župnik '	📄
19	<input type="checkbox"/>	📘	Matica Slovensk... tar Jožef, župnik v Gornji <b>Lendavi</b> . Skrlec Jožef, gostilničar	📄
20	<input type="checkbox"/>	📘	Matica Slovensk... ti F el so Lendva, Gornja <b>Lendava</b> (OberLimbach). Na eni st	📄

Rows per page: 20 ▾ 1-20 of 152    ⏪ ⏩ 1 / 8 ⏪ ⏩

## Rezultat

**149.391 izvodov**  
**objavljenih na**  
**konkordančniku**

<https://www.clarin.si/ske/#concordance?corpname=speriodika>





# Naslednji korak: "Segmentacija zgodovinskih tiskanih medijev z uporabo vizualnih in jezikovnih vložitev"

EUGENIO VENUTI j Novi slovenski ceniki s dvokoles, šival. In kmet. strojev, gramofonov in vsakovrst. plošč, itd. franko. Izvršuje 83 tudi po-Prodajalnica pohištva in manifakturnega blaga K "BatJem. 9»ricaSalutin" Poskusite FI-GOVO KAVO 99 ne j zd r a v s j s i kavni ivni p r i d a t e k . tolišib [jšib p r r o o d d a j a l n i c a h . ki Je naj- finejii Dobiv« m v v s e h p i a z z a S . C a t e r i n a S t . 3 . t r s t , M . German i OKREPLJU-JUCTZELUDFICNI EL1S1R1 Makso Smolenskv (zajamčeni izdelki) toliko p r o t i t a k o j š n j e m u i p l a č i l u k o l i k o r n a o b r o k e . p r o d a j a p i v o v s o d i h I n v s t e k l e n i c a h 0 1

Spedijska in **BALKAN** v Trstu, Riva Grumula 14.  
 - komisijonalna družba  
 Telef. 21-60. Brzjavni: Balkansped.  
 Žiro-račun pri Jadranski banki v Trstu.

Sprejema v svoja dobro urejena skladišča vsakovrstno blago in daje na isto predujme pod najpovoljnejšimi pogoji, daja carinska posojila in ocarinjuje blago, preskrbuje prodajo blaga ter posreduje kupovanje in prodajo vsakovrstnega blaga po nalogu in na račun svojih odjemalcev najkulantneje. Preskrbuje vsakovrstne pošiljkatve na vse kraje hitro in točno.  
 Obavlja vse reklamacije in daje vsakovrstna strokovna pojasnila in navodila takoj in brezplačno.

**Novi slovenski ceniki :-:**  
 dvokoles, šival. In kmet. strojev, gramofonov in vsakovrst. plošč, itd. franko. Izvršuje se tudi po-prave Na: Batjeta, Gorica  
 poskusite FI-GOVO KAVO „Salutin“ ki je naj- finejši in najzdravejši kavni pridatek. Dobiva se v vseh ožjih prodajalnicah.

**EUGENIO VENUTI**  
 Prodajalnica pohištva in manifakturnega blaga  
 (zajamčeni izdelki) toliko proti takojšnjemu plačilu kolikor na obroke.  
 Trst, Piazza S. Caterina št. 3. I. nadstropje.

OKREPLJUJUCI ZELUDACNI ELISIR  
 VLAHO  
 ZELUDACNI ELISIR

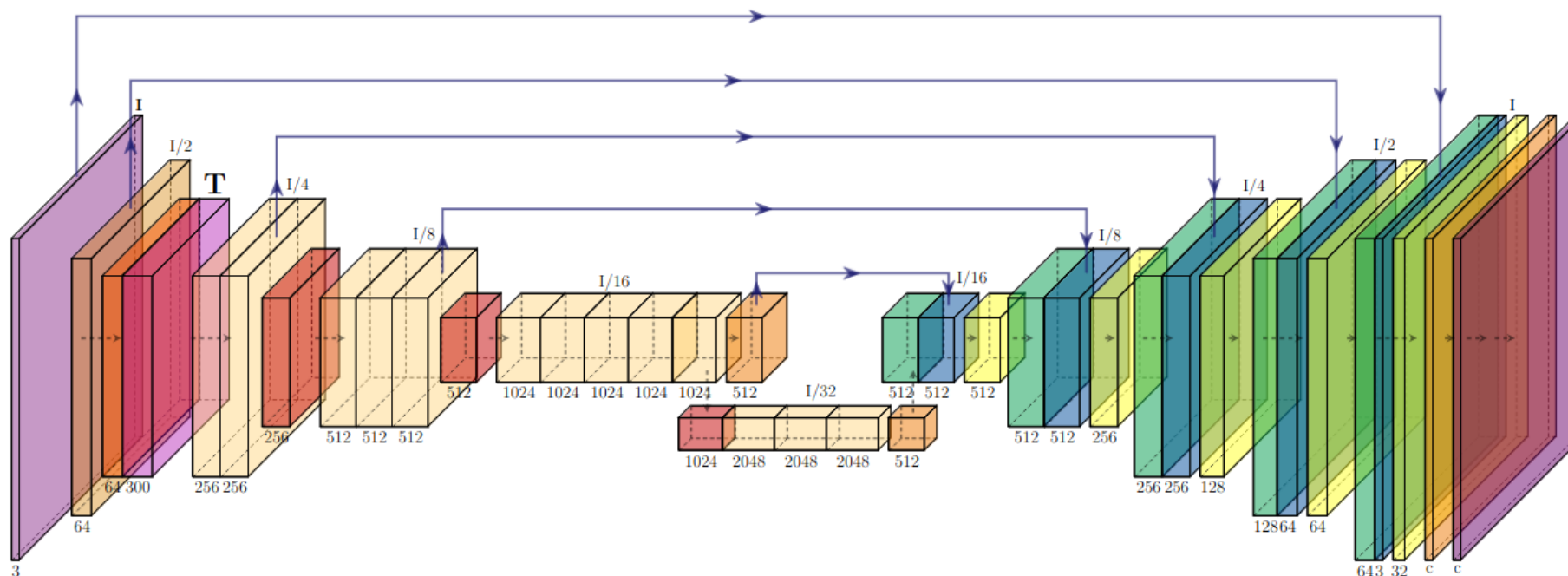
**M. German**  
 Trst, ulica Ponterosso 3, Trst  
 Zaloga semen. Specijaliteta sladkornega régrata in špináče.  
 Na kajo pošljete se ilustrirani ceniki sestoj in pošlino gratis.  
 Prodajalnice ovetilo: Ulica Michelangelo Buonarroti.

**Makso Smolensky**  
 TRST  
 generalno zastopstvo in zaloga večkrat odlikovane pivovarne in (jedmona) tovarne kneza A. J. Schwarzenberga v Prohivnu (Češko) za Trst, Primorje, Istro, Dalmacijo, države sredozemskega morja in dežele ruščega morja.  
**Prodaja pivo v sodih in v steklenicah**  
 avitile Izvazno pivo (á la Pilsen) črno granatko pivo (á la München)  
 Zaloga za Trst in notranje dežele ul. Gielano Donizetti 3, Telef. 2284  
 Izvazna piazras ulica TORRE BIANCA št. 20. Telefona 2315. Pri naroči v od najmanj 12 steklenic 2000 prostu vročitve na dom.

# Pogled v prihodnost

"[...] the targeted image segment typology goes beyond physical and/or logical characteristics and considers fine-grained semantic content item types (e.g. a segment is not only an article, but also e.g. a serial or death notice, or not only a table, but also e.g. election results or stock exchange information)".

Barman et al. F.: Combining visual and textual features for semantic segmentation of historical newspapers.



- Input image
- Concat embeddings feature map
- Convolution 7x7 + maxpool 2x2
- Maxpool 2x2
- Resnet bottleneck
- Resnet bottleneck + maxpool 2x2
- Convolution 1x1
- Bilinear upsampling 2x2
- Concat
- Convolution 3x3
- Softmax
- Copy



# Hvala za pozornost!

- Kontakt in nekaj povezav:

[filip.dobranic@inz.si](mailto:filip.dobranic@inz.si)

<https://www.inz.si>

<https://dihur.si>

<https://clarin.si>

**Meteorološko poročilo.**

Dan	Čas opazovanja	Stanje barometra v mm.	Temperatura	Vetrovi	Nebo	Mokrina v mm.
5. okt.	7. zjutraj	731.7 mm.	11.4° C	sl. zah. obl.		6-10 mm
	2. popol.	733.0 mm.	18.0° C	sl. jzh. obl.		
	9. zvečer	734.9 mm.	15.6° C	sl. zah. d. jas.		dožja.
Srednja temperatura 15.0°, za 2.0° nad normalom.						

**Dr. IVAN MILAN HRIBAR**  
 vljudno naznanja, da je  
**otvoril**  
 svojo odvetniško pisarno  
 v Ljubljani  
 Sodnijske ulice števil. 2.

**Radenska**  
 kislina voda po natriju in litiju najbogatejša



preskušan. o zdravilno sredstvo proti mehurnim boleznim, kamenu v mehurju, protinu, dalje proti boleznim v zledenci, mokrih, dolgotrajnemu kataru dihal, zlatej žili in zlatenici.

Poskusni doktorjev: Garrod, Biswanger, Cantani. Ure so dokazali, da ima ogljenokislina litij največjo raztopilno moč pri sečnokislinskih usedlinih izločkih, iz česar se sklepa na najugotovljivejšo učinko Radenske kisline vode.

Kot okrepilna pijača z vinom ali sadnimi sokovi in sladkorjem pomešana, je Radenska voda v obče priljubljena.

**Kopališče Radensko**  
 Kopanje v slatini in v jeklenici upliva posebno pri: protinu, bolečinah mokrih, malokrvnosti, ženskih boleznih, slahostih i. t. d.

**Stanovanjske počitnice**  
 Prospekt (brezplačno od ravateljstva: kopališče Radensko blizu Ljutomeru)

**Radenska kislina voda, ob vznožju Slovenskih gor, ne zamenjati z Radgonsko, to je Radkersburger.**

Čeravno Radenska kislina voda stoji več kot jeden dan v odprti posodi, vendar se peni prav močno, ako se z vinom pomeša, ker ima v sebi spojene ogljikove kisline. To svojstvo pa jo odlikuje pred mnogimi drugimi kislimi vodami, ki se že pogubi in razvada le prosto ali manj trdno spojeno ogljikovo kislino, ki se že pogubi in razvada, kakor hitro se steklenica odmaši. (264-5)

Prvega petelinca v letošnji sezoni je v soboto 1. aprila ustrelil na Kolescu ljubljanski zobozdravnik g. dr. Bretl.

**Poslano.\***

Res je, da sem svojo ženo A. R. pretepel zaradi njenega hudega jezika in ker namesto mene ljubi „repetničarje“. Če bi ji bil jaz rebro zlomil, ne bi bila v štirih dneh zdrava.

**Anton Rajk,**  
 Vavpčavas, Dobernoče, Dolenjsko.

\*) Za vsebino tega spisa je uredništvo odgovorno le toliko, kolikor določa zakon. (2679)



# Financiranje

Projekt EuroCC 2 financira Evropska unija. Financiran je s sredstvi Skupnega podjetja za evropsko visokozmogljivo računalništvo (EuroHPC JU) ter Nemčije, Bolgarije, Avstrije, Hrvaške, Cipra, Češke republike, Danske, Estonije, Finske, Grčije, Madžarske, Irske, Italije, Litve, Latvije, Poljske, Portugalske, Romunije, Slovenije, Španije, Švedske, Francije, Nizozemske, Belgije, Luksemburga, Slovaške, Norveške, Turčije, Republike Severne Makedonije, Islandije, Črne gore in Srbije v okviru sporazuma o dodelitvi sredstev št. 101101903.

Delovanje Nacionalnega kompetenčnega centra SLING sofinancira Ministrstvo za visoko šolstvo, znanost in inovacije.

## Medijski sponzor

**Računalniške  
novice**

[www.racunalniske-novice.com](http://www.racunalniske-novice.com)



**Co-funded by  
the European Union**



**REPUBLIKA SLOVENIJA  
MINISTRSTVO ZA VISOKO ŠOLSTVO,  
ZNANOST IN INOVACIJE**