

Jauninājumi datu apstrādes un pētniecības infrastruktūrā LNB: bibliogrāfisko datu un mākslīgā intelekta fokuss

Anda Baklāne, Matīss Bolšteins, Marta Kivkule, Eduards Skvireckis

Latvijas Nacionālā bibliotēka

CLARIN & DARIAH pavasara konference, 05.03.2026.

Kontakti: anda.baklane@lnb.lv, dh@lnb.lv

Mākslīgā intelekta lietojumgadījumi LNB

- Satura rādītāju izveide video digitalizācijā
- Teksta datu kopu izveide, klasifikācija
- Jēdzienu izrace (*concept mining*)
- Optiski atpazīta teksta kvalitāte, normalizācija
- Attēlu grupēšana, klasificēšana, atpazīšana
- Bibliogrāfisko ierakstu ģenerēšana un papildināšana
- MI aģentu izmantojums atsauces datu sistēmā

Satura rādītāju izveide

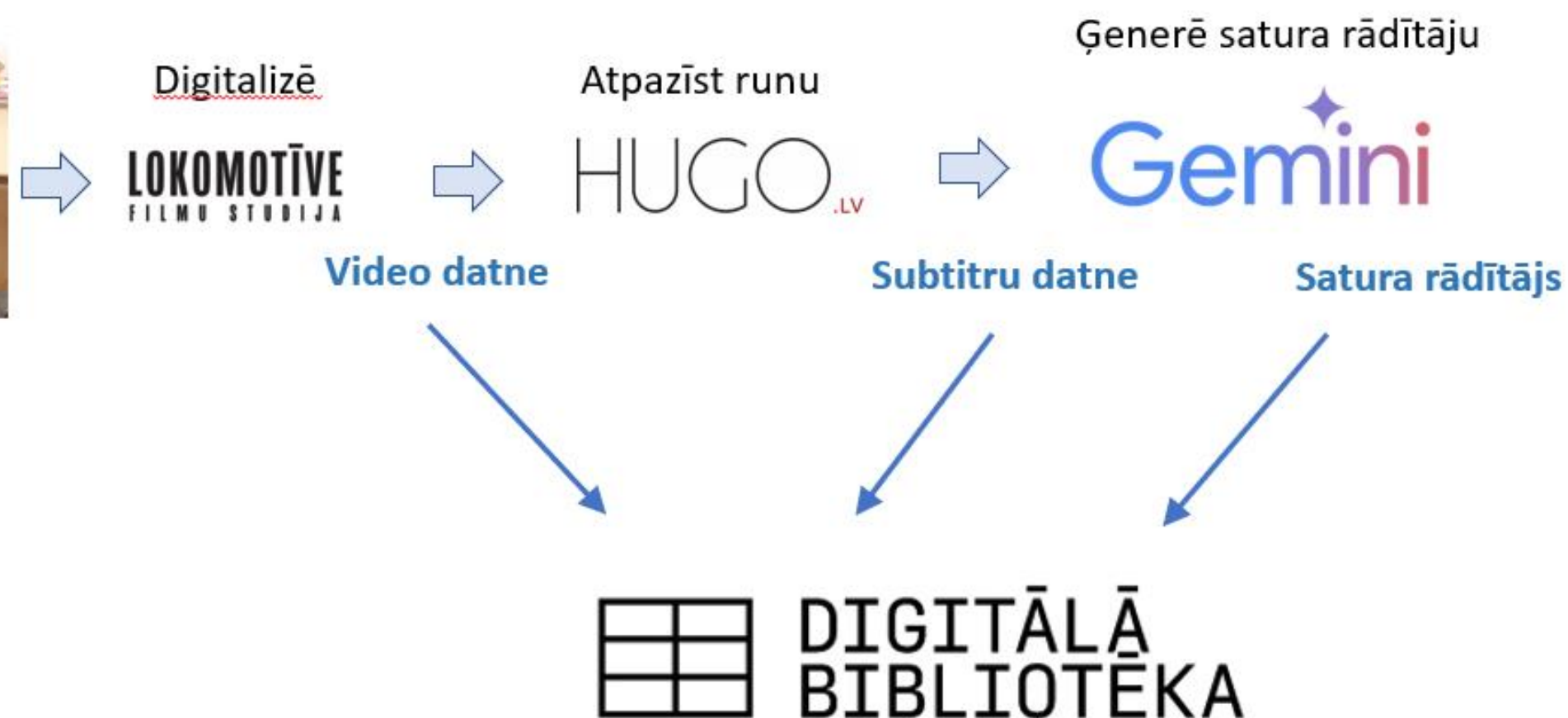
No atpazītas runas izveidoti satura rādītāji - video ierakstu sadalījums sižetos ar laika norādēm

Sākotnējais apjoms: 2000 stundas
Risinājums: *Gemini 2.0*
Skaties [Digitālajā bibliotēkā](#)

ŠIS IR MĀKSLĪGĀ INTELEKTA VEIDOTS KOPSAVILKUMS, UN TAJĀ VAR BŪT KĻŪDAS VAI NEPRECIZITĀTES. Ievads un sveiciens (00:00:03); Saruna par deputātu (00:03:30); Pasākuma ievads un aicinājums piedalīties (00:04:50); Saruna par ēdienu un plāniem (00:06:17); Saruna par notikuma programmu (00:06:56); Mūzikas un mākoņu aizdzīšanas tēma (00:07:12); Saruna par līgumu un zināšanām (00:11:37); Saruna par bērniņu un kino (00:15:28); Saruna par teātra vai zāles telpām (00:16:57); Dzejas lasījums par vīrieti (00:19:44); Saruna par



Reģionu TV saturs:
5 750 stundas



Mūzikas tekstu korpusa izveide (1864–2022)

- teksti par mūziku, mūziķiem, koncertiem, ierakstiem
- 115 laikrakstu nosaukumi, ap 160,000 raksti
- sarežģīta tekstu izguve, daudz kļūdaino atbilstmju, balstoties atslēgvārdos

LVM izmantojums

- *Gemini Flash 1.5* demonstrēja pārlicinošus rezultātus, 100% preciz. testa datos
- Apšaubāmi gandrīz tikai tie rezultāti, kas arī subjektīvi neviennozīmīgi
- Plašas iespējas automatizēti klasificēt rakstus un piešķirt birkas



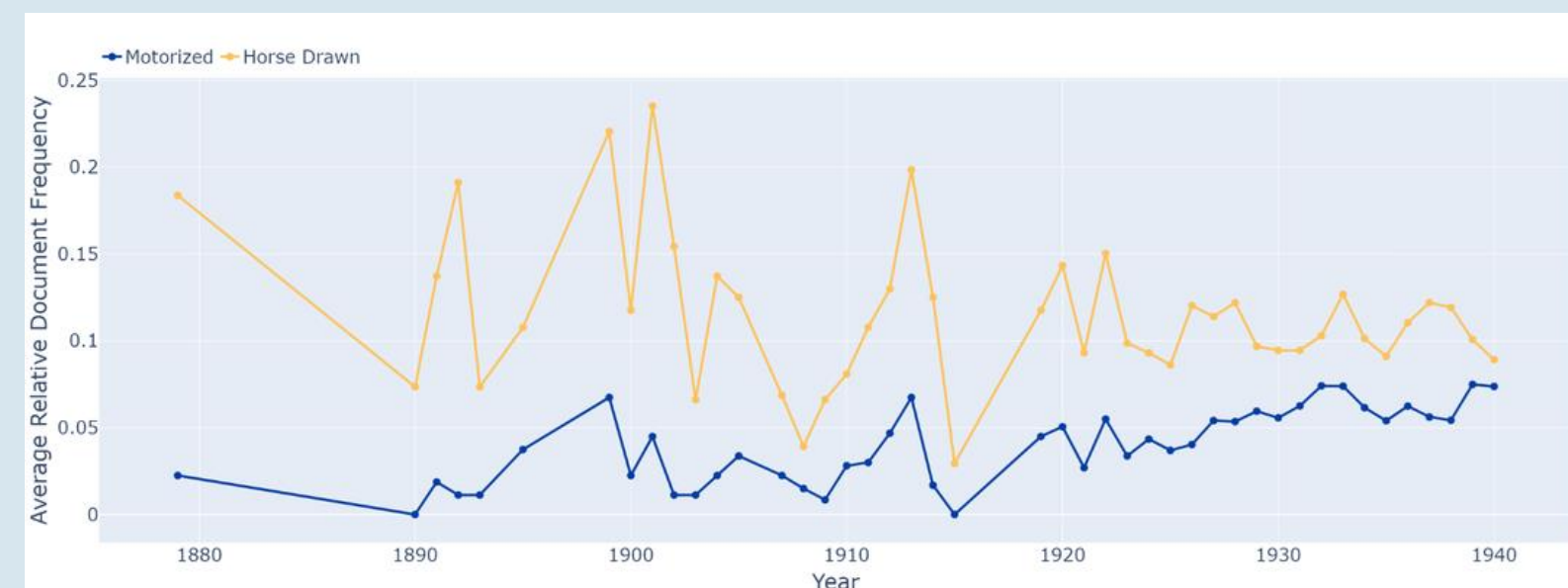
Transports latviešu senākajos romānos - jēdzienu izrace



Latviešu senākie romāni (1879–1940), 463 darbi

<https://digitalbiblioteka.lv/?col=67d84fb950b39>

<https://dom.lndb.lv/data/obj/1554666>



jcls. Journal of
Computational
Literary Studies

OPEN  ACCESS

under review

Urban Transportation in the Latvian Early Novels or "Why do you use a 19th-century horse-drawn cab when you have a 20th-century taxi?"

Eva Kristsone-Eglāja¹ 

Anda Baklāne²

Valdis Saulespurēns²

<https://doi.org/10.26083/tuprints-00030146>

Vecās drukas tekstu kvalitātes uzlabošana un normalizācija

title: Dazchadas ziņas

subheadline:

author:

section:

uri: <http://dom.lndb.lv/data/obj/580349>

Par Domes darheem .·pehz

- »· ftp-Übert» Domes preet chneezcba 8.

; dez. spreaduse par to, ko pe z fwehtkeem

J darit Wispirms nemfchot preekschå bu-

»dschetu, pehz kam liku»m·a projektu par

«. Holmas gubernas nodlblnafchcmu, tad

A otrreik apfpreedischot likumu par kaxa

kkausi u un eerehdnu atbildibu un strahd"

neeku likumusl

L No ~ihfto kreewu kaufchu« no-

- J flehpumeew

s Tumschi. noflehpumaini dccrbi, tumfcha

title: Dažādas ziņas

subheadline:

author:

section:

uri: <http://dom.lndb.lv/data/obj/580349>

Par Domes darbiem pēc

svētkiem. Domes priekšniecība 8.

decembrī spriedusi par to, ko pēc svētkiem

darīt. Vispirms ņemšot priekšā budžetu,

pēc tam likuma projektu par Holmas

gubernas nodibināšanu, tad otru reizi

apspriedišot likumu par kara klausību un

ierēdņu atbildību un strādnieku likumus.

No „īsto krievu kaušu” noflehpumiem.

Tumši, noslēpumaini darbi, tum-

ša nauda — to vien dzird rakstam par



Sākotnējais

Gemini-2.5-flash-preview

CNN balstīta attēlu grupēšana kolekciju pārlūkošanai un pētniecībai

<https://zenta.lnb.lv/>

Z. Dzividzinskas arhīvs

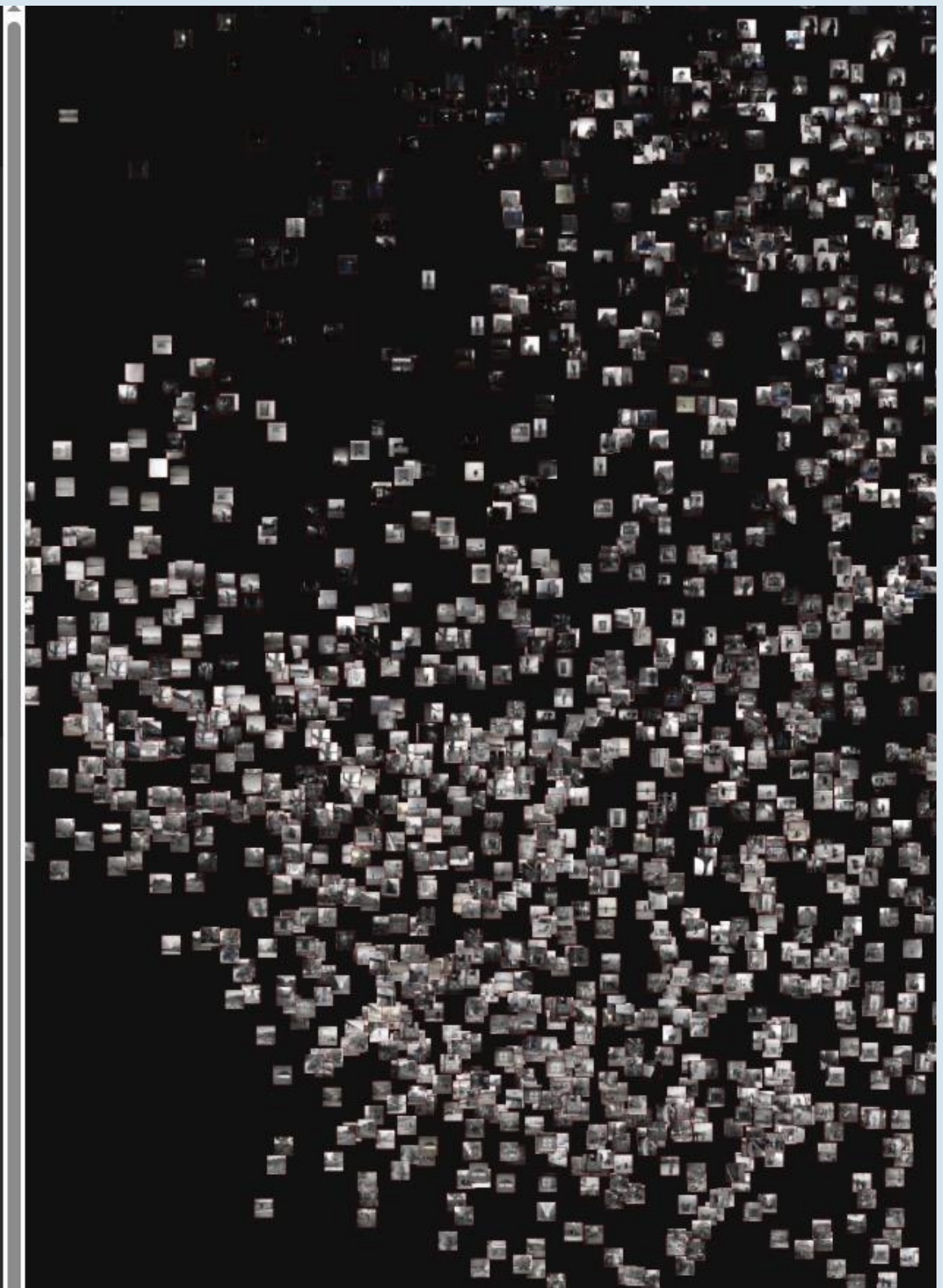
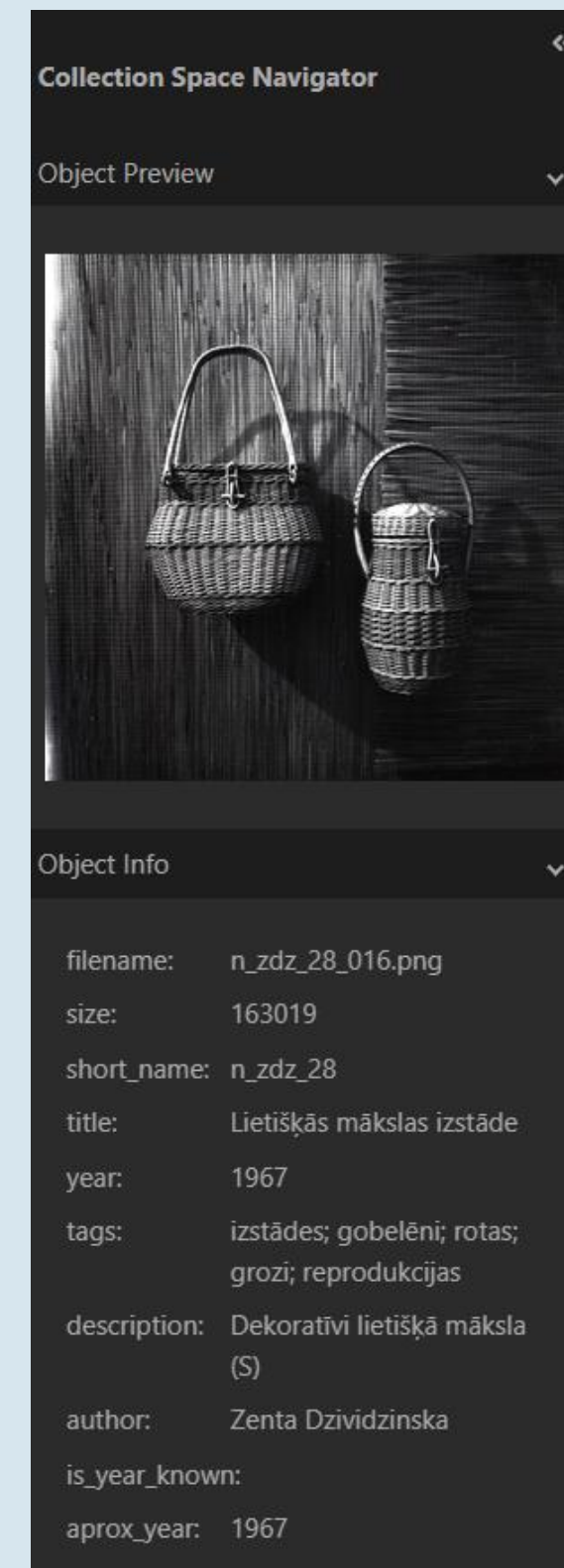
<https://digitalabiblioteka.lv/?col=1549393>

Balstīts:

<https://collection-space-navigator.github.io/>

Prototipa risinājumā izmantots ResNet-50 attēlu modelis; mākonis - izmantojot PCA algoritmu.

Atslēgvārdu piešķiršana attēliem, papildu apstrāde: *GPT-4o*



CNN balstīta attēlu grupēšana kolekciju pārlūkošanai un pētniecībai

<https://zenta.lnb.lv/clip>
<https://zenta.lnb.lv/clip/multi>
Multimodāls modelis CLIP (attēls + teksts)

Results for "sunny day"

Enter your search...

Top N results: 20

Search

Download CSV



n_zdz_1965_112_2.jpg — 0.249



n_zdz_1965_111_5.jpg — 0.244



n_zdz_56_005.jpg — 0.243



n_zdz_83_002.jpg — 0.243



n_zdz_83_006.jpg — 0.242



n_zdz_1965_550_1.jpg — 0.241



n_zdz_1965_81_5.jpg — 0.241



n_zdz_82_007.jpg — 0.240



n_zdz_56_002.jpg — 0.239



n_zdz_82_006.jpg — 0.239

Radar (Top matches)

Aggregate confidence by query (non-zero + scaled)
Green = non-zero ratio (count/N). Blue = scaled confidence ratio (sum of non-zero confidences / N).

Toggle Labels: Concepts

Toggle Non-Zero: On

Toggle Scaled: On

Download SVG

a high-angle shot

a Dutch-angle shot

an eye-level shot

a low-angle shot

1. a high-angle shot (sum=0.681+0.0000000 scaled=0.571+0.0000000)

2. an eye-level shot (sum=0.388+0.1110000 scaled=0.241+0.0000000)

3. a low-angle shot (sum=0.766+0.0000000 scaled=0.388+0.0000000)

4. a Dutch-angle shot (sum=0.000+0.0000000 scaled=0.000+0.0000000)

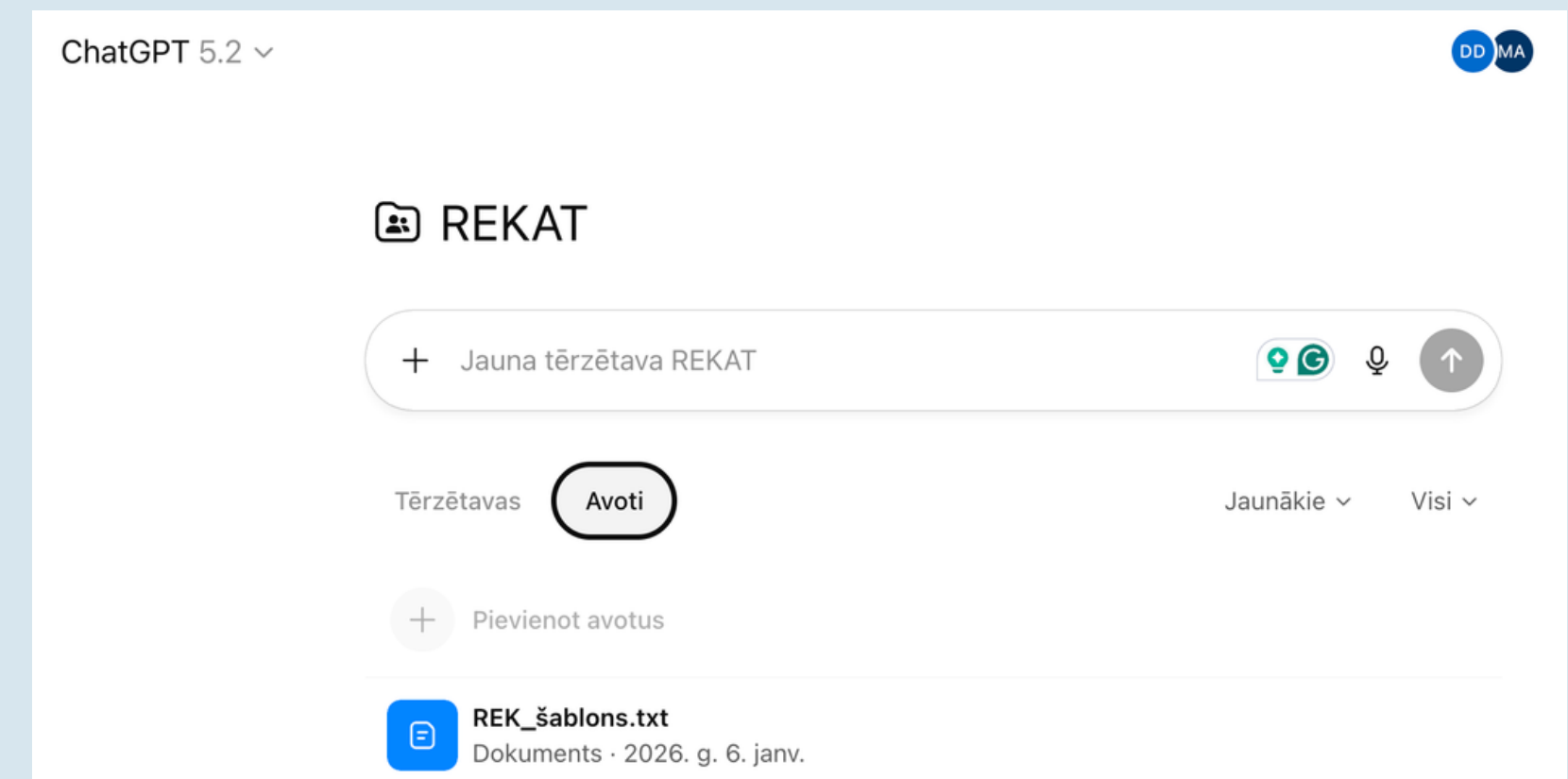
Bibliogrāfisko ierakstu ģenerēšana

Mērķi

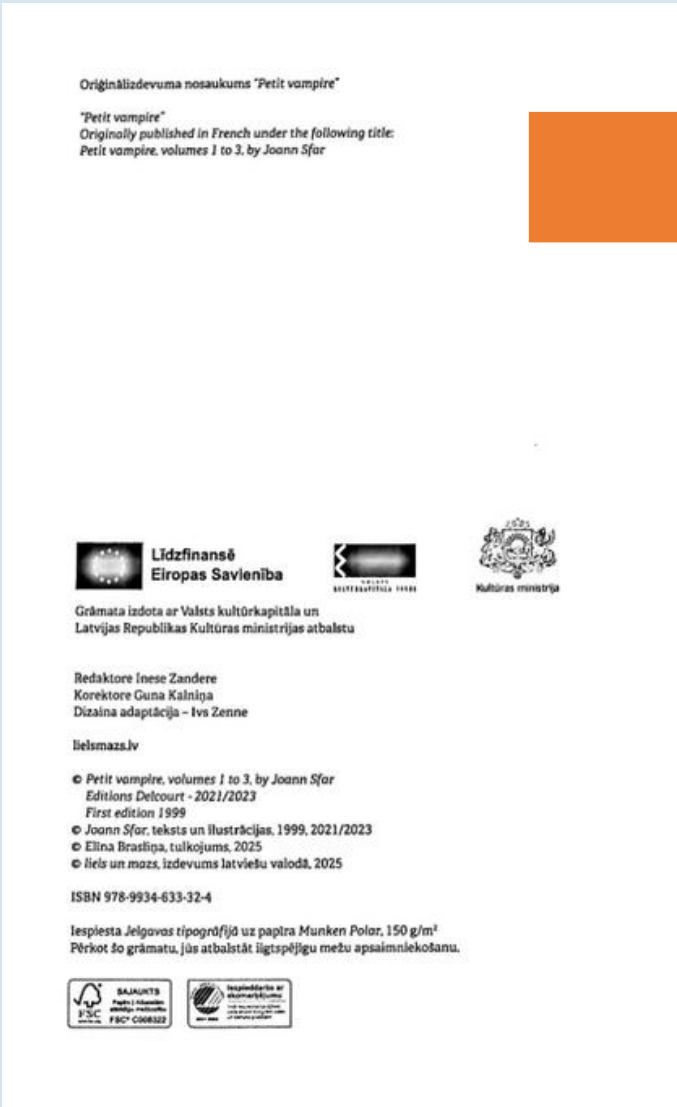
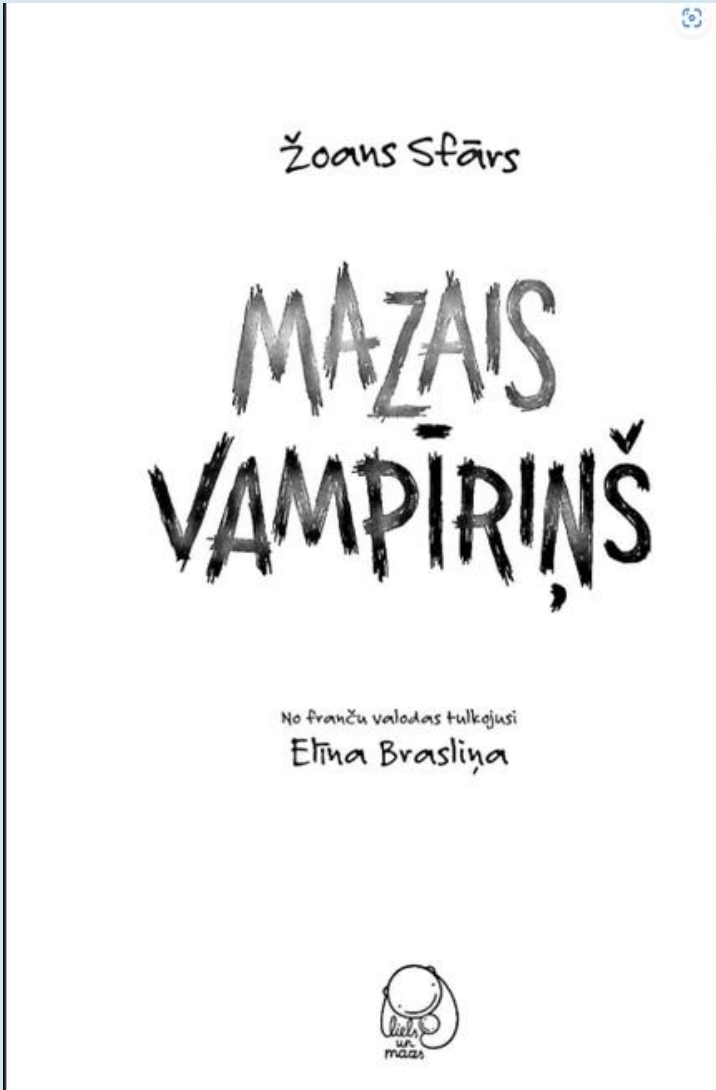
- Retrospektīvā katalogizācija
- Atbalsts publiskajām bibliotēkām

Ko darām?

- Izmantojam ChatGPT projektu
- **Ievietojam titullapu attēlus**
- **Izgūstam strukturētus bibliogrāfiskos aprakstus**
- Balstāmies uz iepriekš definētu šablonu



Bibliogrāfisko ierakstu ģenerēšana



Šeit ir MARC21 ieraksts grāmatai "Mazais vampīriņš":

pgsql Copy Edit

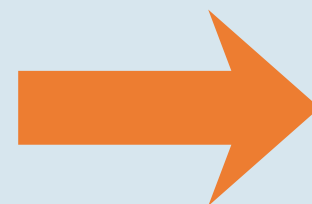
```
000000001 FMT L BK
000000001 LDR L ^^^^^nam^a22^^^^^^i^4500
000000001 008 L 250613s2025^^^^ly^^^^^^^^^^^^000^0^lav^^
000000001 020 L $$a9789934633324$$qbrošēts
000000001 040 L $$aLV-RiVB$$blav$$enda
000000001 041 L $$alav$$hfre
000000001 100 L $$aSfārs, Žoans.$$4aut
000000001 240 L $$aPetit vampire.$$1Latviešu valodā
000000001 24510 L $$aMazais vampīriņš /$$cŽoans Sfārs ; no franču valodas tulkojusi Elīna Brasliņa
000000001 264 1 L $$a[Rīga] :$$bliels un mazs,$$c[2025]
000000001 264 3 L $$a[Jelgava] :$$bJelgavas tipogrāfija
000000001 300 L $$a101 lpp. :$$bilustrācijas ;$$c30 cm
000000001 336 L $$ateksts$$btxt
000000001 336 L $$aattēls$$bsti
000000001 337 L $$atiešuztvere$$bn
000000001 338 L $$a$ējums$$bnc
000000001 590 L $$aIeraksta pamatlīmenis izveidots ar mākslīgā intelekta rīka palīdzību.
000000001 7201 L $$aBrasliņa, Elīna.$$4trl
000000001 7201 L $$aZandere, Inese.$$4edt
000000001 7201 L $$aKalniņa, Guna.$$4edc
000000001 7201 L $$aZenne, Ivs.$$4bkd
000000001 7102 L $$aLiels un mazs.$$4pbl
000000001 91500 L $$a2025.NLL
000000001 91503 L $$ak20250613.03
000000001 91503 L $$a202506.rgb
000000001 919 L $$anb
```

Bibliogrāfisko ierakstu ģenerēšana

- **Atvieglo metadatu ievadi** un ietaupa laiku
- Īpaši resursiem svešvalodās un vecajā ortogrāfijā (piemēram, krievu, vācu fraktūra)
- Apraksta **pārbaude ir obligāta** – MI nevar pilnībā aizvietot katalogizētāja darbu
- Iespējami **neparedzami rezultāti** – halucinācijas

Nākotnei:

- pētām pakešapstrādes (*batch processing*) iespējas



Valoda: English Latvian

Pakešu apstrādātājs

Attēlu pārveidotājs MARC tekstā

Masveida apstrādes rīks attēlu pārveidošanai MARC tekstā

API atslēga:

Sistēmas uzvedne:

Ielādēt no static/prompts/marc.txt

Lietotājsvārds:

piem., rek_user

Pielāgots nobeigums:

```
000000001 91500 L $$a2026.NLL
000000001 91503 L $$arek20260305.03
```

Izvēlieties modeli:

Gemini 3 Flash Preview (noklusējums)

Vai ievadiet pielāgotu modeļa ID:

piem., openai/gpt-4.1-mini

Augšupielādēt ZIP (nav obligāti, ja zemāk izvēlaties mapi):

Choose File No file chosen

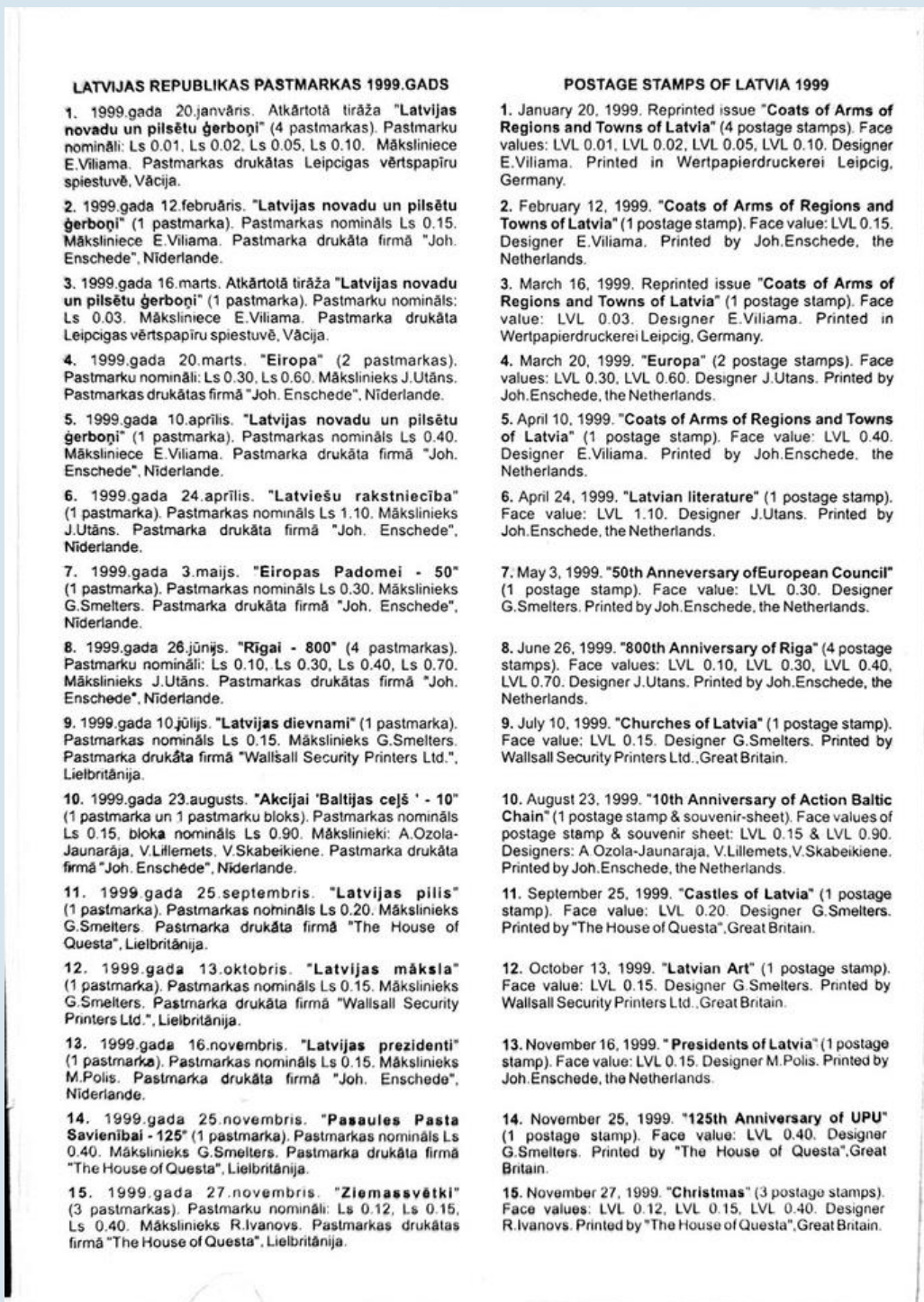
Vai izvēlieties apakšmapi no /mnt/mi_rek:

Marta/	13 vienības 2026-02-20 14:49:18
tttt/	6 vienības 2026-02-20 14:48:56
testamape/	11 vienības 2026-02-18 15:31:14
Di?na/	0 vienības 2026-01-20 13:59:06

☐ Iekļaut ievades failus rezultātu zip arhīvā
☒ Atdalīt teksta failus rezultātu zip arhīvā
☐ Iekļaut metadata.json
☒ Saglabāt apvienotos rezultātus šeit /mnt/mi_rek/results

Iesniegt darbu

Bibliogrāfisko ierakstu papildināšana



Pastmarku katalogi (1997-2024)

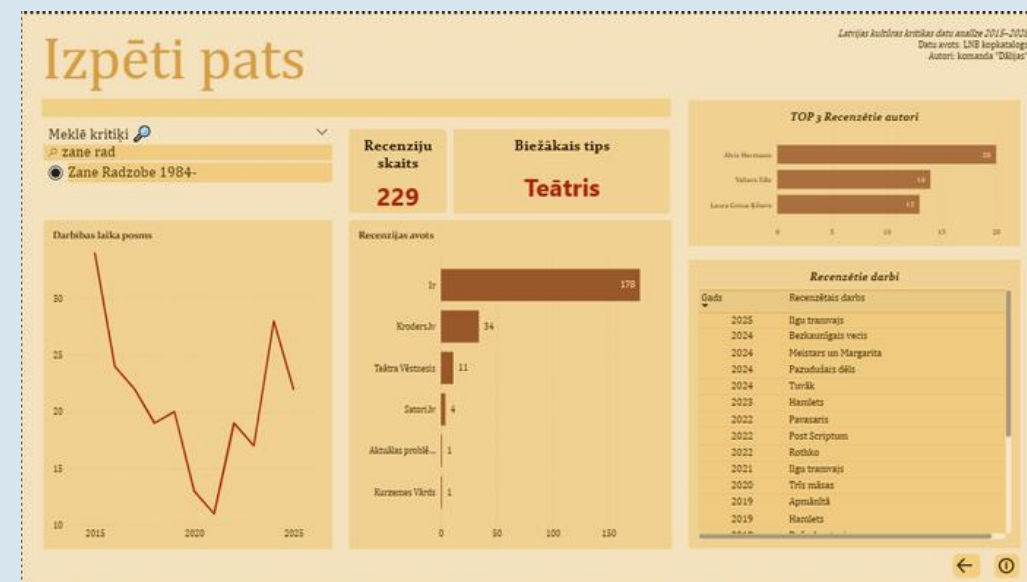
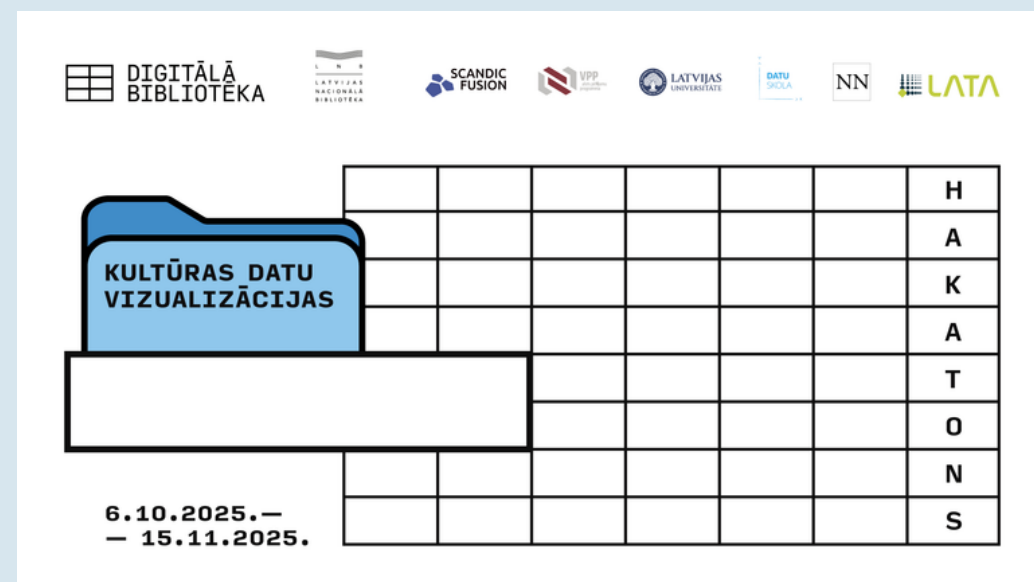
Papildināti 32 ieraksti ar 904 laukiem

Kultūras datu vizualizācijas hakatons (2025)

- bibliogrāfisko datu kopa: teātra, kino, literatūras, mākslas, mūzikas recenziju analītikas dati
- digitalabiblioteka.lv metadati

Visi konkursa projekti

Projekts Mākslas kritikas Zelta balva





ATS — Atsauces datu sistēma

Latvijas kultūras mantojuma zināšanu organizēšanas sistēma · VIS statuss (MK 23-TA-485) · DKM apakšsistēma

ENTĪTES

Tips	
Persona	~
Temats	
Institūcija	
Darbs	
Vieta	
Notikums / Forma-Žanrs / Grupa / Dzimta	

659 000 attiecības (150+ tipi)

91% entīšu "uzticams"

INTEGRĀCIJAS

ATS

Aleph

NMKK

DOM

LNA IS

APLIS

DZSK

REST API (240+) · iFrame + JWT

ĀRĒJIE IDENTIFIKATORI

~1 000 000

saišu · 25+ sistēmas

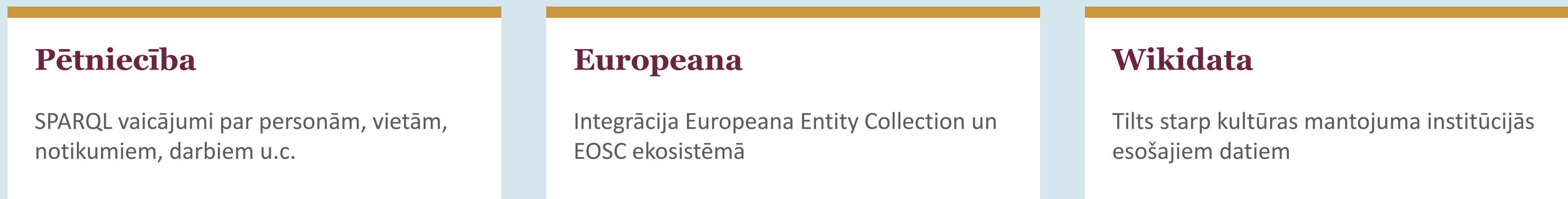
Aleph	360k
VIAF	245k
ISNI	166k
Wikidata	100k
LCSH	43k
FAST	13k
YSO	10k
ULAN	5,4k

+ ORCID · GND · LCNAF · LĢIA · filmas.lv · u.c.

Semantiskā tīmekļa vīzija



Ko tas atver:



MI vīzija — *ATS Toolkit*

Python SDK · MCP serveris · MI aģenti

IEKŠĒJĀ LIETOŠANA — datu pārvaldība

Entīšu saskaņošana	ATS ↔ Aleph, DOM, DZSK — saskaņošana starp sistēmām
Datu bagātināšana	Automātiska Wikidata, VIAF, ISNI saišu pievienošana
Dublikātu atklāšana	Semantiskas līdzības meklēšana 504k+ entītēs
Kvalitātes novērtēšana	Nepilnīgu, pretrunīgu ierakstu skenēšana
Vārdformu normalizēšana	Vēsturisku vārdu variantu standartizēšana
Attiecību klasifikācija	67% «nenoteikto» attiecību pārkvalificēšana

ĀRĒJĀ LIETOŠANA — pētniekiem

Aģentiskie darba procesi

Toolkit ļauj pētniekiem analizēt ATS datus ar MI aģentu palīdzību:

Prosopgrāfija

354k personu biogrāfiskā analīze

Tīklu analīze

659k attiecību grafu izpēte

Ģeogrāfiskā analīze

65k dzimšanas / 19,5k dzīvesvietu sasaiste

Tematikas izpēte

60k tematu un 2,9k žanru klasifikācijai

Dabiskās valodas vaicājumi

Jautā par ATS datiem — saņem atbildes

MI mērogā: 504k entītes · 659k attiecības · 1M ārējo saišu — pārāk daudz cilvēkam, ideāli MI aģentam

KPLIIA

Koplietošanas platforma lielapjoma informācijas izguvei un analīzei

MĒRĶIS

Izveidot koplietošanas informācijas sistēmu, kas nodrošinās automatizētus pakalpojumus liela apjoma tekstu un audiovizuālās informācijas izgūšanai no publiskiem tiešsaistes resursiem, satura importu un iegūtā informācijas masīva kvalitatīvu analīzi.

PIELIETOJUMS

- Dezinformācijas un propagandas atklāšana
- Spriedzes eskalācijas agrīnā brīdināšana
- Rīcībpolitikas veidošanas atbalsts

PLĀNOTIE PAKALPOJUMI



Datu rasmošana, savākšana un attīrīšana



Teksta transkribēšana (speech-to-text)

Hugo



Teksta tulkošana

Hugo



Sentimenta analīze

LUMI / HuggingFace



Entīšu atpazīšana (NER+)

PiniTree / LUMI



Tekstu līdzības un dedublēšanas analīze

Minhash



Valodas noteikšana

Paldies!

Atbalsta:

Projekts "Latvijas kultūras ekosistēma kā resurss valsts izturētspējai un ilgtspējai"/ CERS, Nr. VPP-MM-LKRVA-2023/1-0001.

Projekts "Atvērtas un FAIR principiēm atbilstīgas digitālo humanitāro zinātņu ekosistēmas attīstība Latvijā (DHELI, 2022–2025)", Nr. VPP-IZM-DH-2022/1-0002.

Projekts "Mediju satura integrācija Digitālā kultūras mantojuma platformā", Nr. 2.1.1.1.i.0/1/23/I/VARAM/007.

Projekts "Koplietošanas platforma lielapjoma informācijas izguvei un analīzei", Nr. 2.1.2.1.i.0/1/24/I/VARAM/012.