

# Tekstizrace bibliotēku digitalizēto resursu kontekstā

Latvijas bibliotekāru 20. konference  
“Bibliotēkas – dialogs ceļā uz pārmaiņām”

16.04.2019

**Anda Baklāne**

anda.baklane@lnb.lv

**Latvijas  
Nacionālā  
bibliotēka**  
*simts*

Bibliotēkas pirmā vadītāja Jāņa Misiņa rokraksts

# Digitālā pētniecība - jēdziens

**Digitālā pētniecība** ir digitālu avotu, digitālu pētniecības metožu, digitālas publicēšanas un saglabāšanas risinājumu izmantojums pētniecības mērķu sasniegšanai.

Saistīti jēdzieni:

**digitālās humanitārās zinātnes** (*digital humanities*)

**digitālās sociālās zinātnes** (*digital social sciences* )

Šaurāki jēdzieni, piemēram:

digitālā vēsture

digitālā stilistika

digitālā žurnālistika

# Datizrace, tekstizrace (*Data mining, text mining*)

Ar **datizraci** saprot

lielu datu kopu analīzi, kas tiek veikta ar mērķi atrast neparedzētas attiecības, kā arī, lai apkopotu datus jaunos veidos tā, lai šo datu kopu analīzes rezultāti datu īpašniekam būtu saprotami un noderīgi.

A. Kapenieks, R. Strazdiņa,  
«Pētījums Nr.1.16  
Datu avotu analīzes un sasaistes rīks»

## **Datizrace**

Datu bāzes datu analīze, kas paredzēta apslēptu sakarību meklēšanai kādā datu grupā [..]

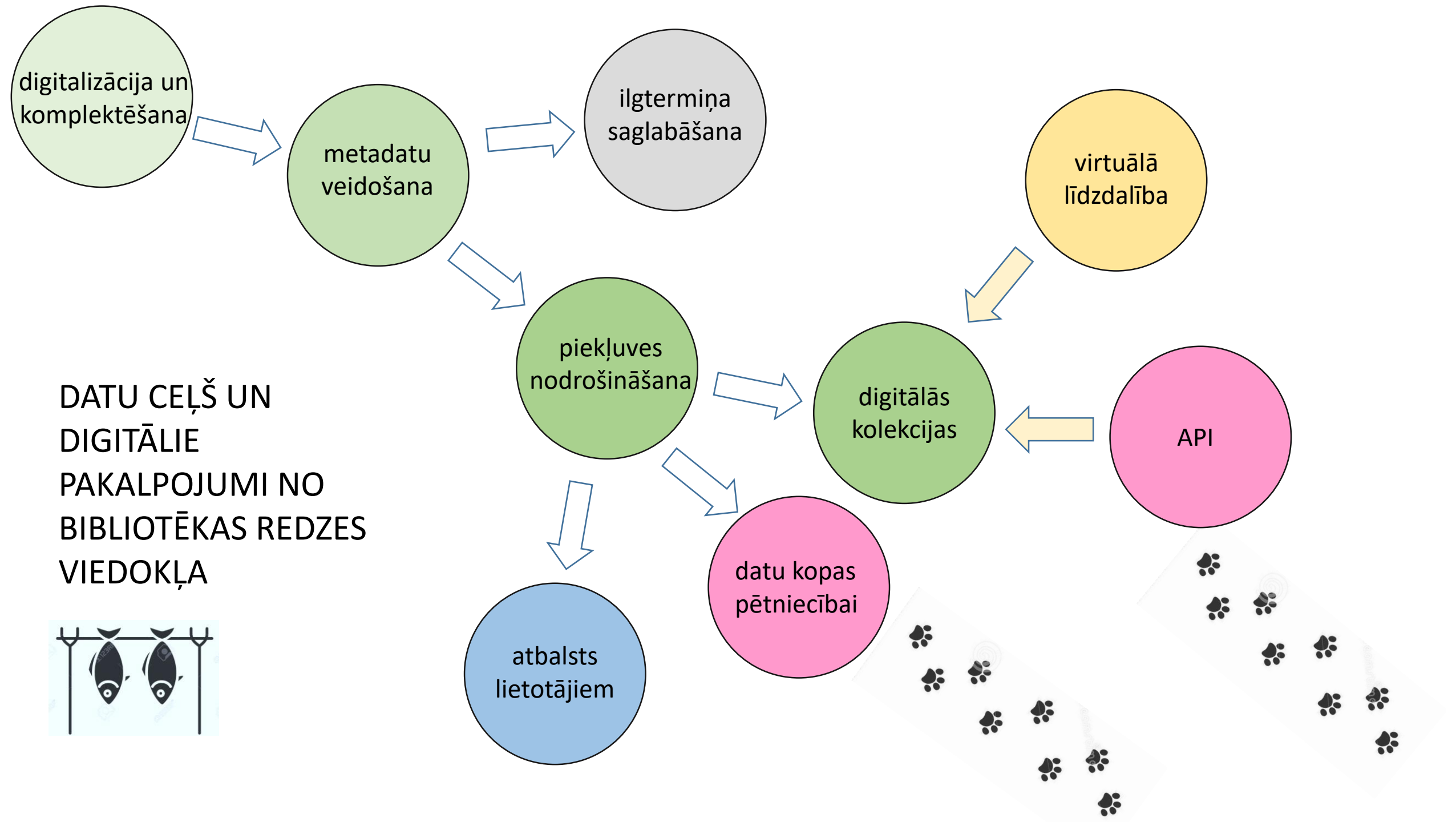
Datizraces programmatūra nemaina datu attēlojumu, bet atklāj iepriekš nezināmas attiecības starp tiem. LZA TK ITTEA terminu datubāze

**Tekstizrace**, savukārt, ir informācijas izguve no nestrukturēta teksta

A. Kapenieks, R. Strazdiņa,  
«Pētījums Nr.1.16  
Datu avotu analīzes un sasaistes rīks»

**Tekstizrace - sakarību meklēšana lielos teksta masīvos**

Bibliotēkas un tekstizrace



## 1. līmenis: piekļuve un informācija

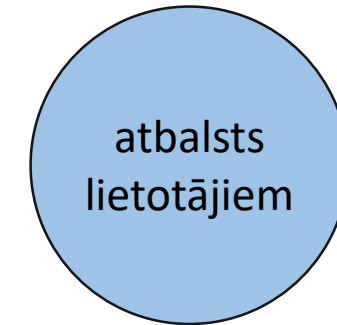
- Lietotājam ir pieejamas digitālas kolekcijas un datu kopas.
- Pieejama informācija par šīm kolekcijām – kādi materiāli tajās atrodami, kādas darbības ar tiem iespējams un atļauts veikt.
- Ir iespēja saņemt konsultāciju par piekļuvi šiem datiem klātienē un attālināti.

## 2. līmenis: atbalsts lietotājiem un standarta risinājumi

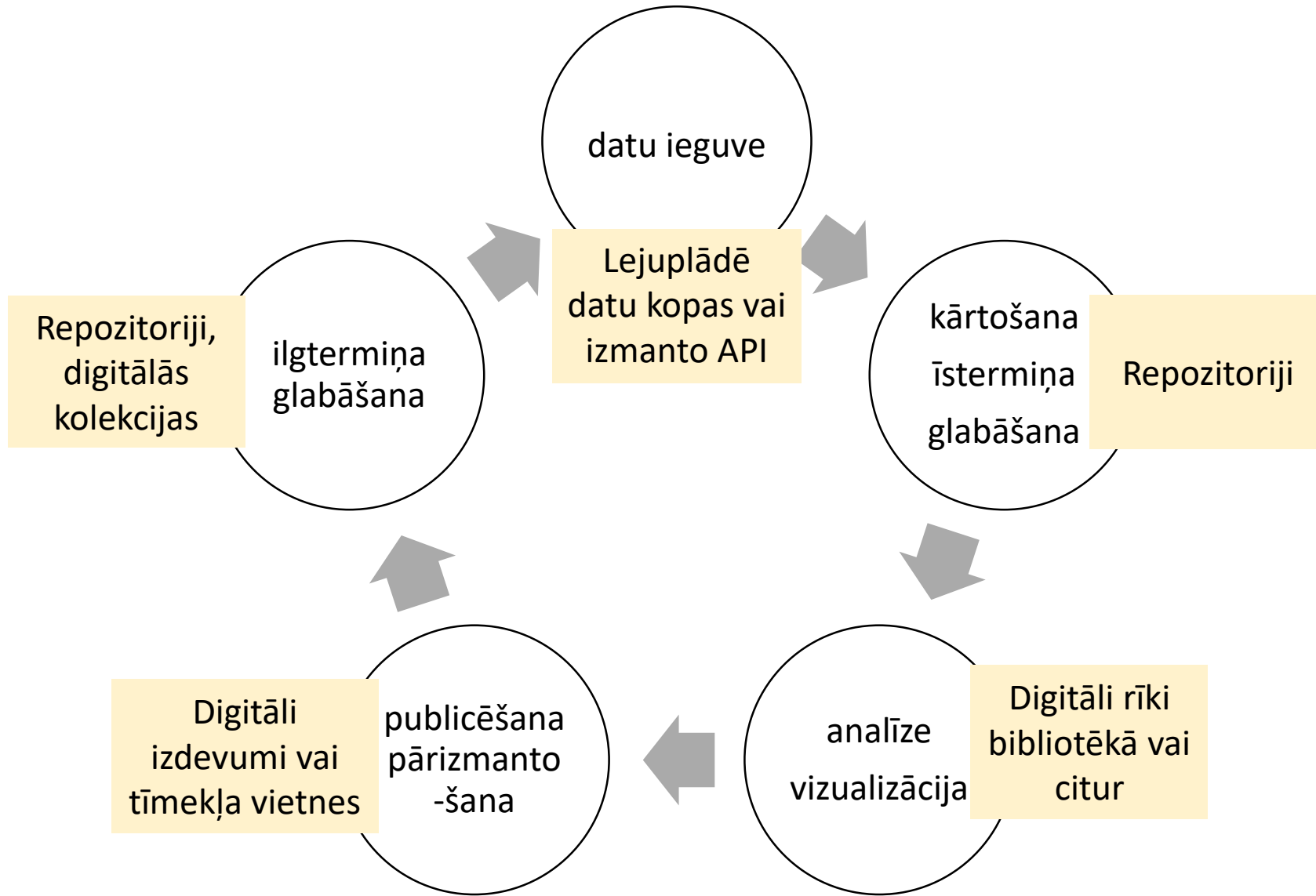
- Bibliotēkas darbinieki var sniegt ieteikumus un palīdzēt izstrādāt digitāla pētniecības projekta metodoloģiju, ja iecere nav sarežģīta.
- Pieejamas lietojumprogrammu saskarnes (API) bibliotēkas datu iegūšanai, daži vienkārši rīki datu analīzei un vizualizācijai.
- Pieejamas pamācības datu lietošanai un īstenotu projektu piemēri, tiek organizētas mācības.

## 3. līmenis: digitālās pētniecības un inovāciju laboratorija

- Laboratorija pastāvīgi izstrādā jaunus rīkus digitālajai pētniecībai.
- Tiek veidoti īpaši individuāli risinājumi pētniecības projektiem, tiek īstenotas rezidējošo pētnieku programmas un projektu konkursi.
- Laboratorijas veidotie novatoriskie risinājumi ar laiku var tikt iestrādāti kā bibliotēkas pamata pakalpojumi.



# DATU DZĪVES CIKLS NO PĒTNIEKA REDZES VIEDOKĻA



Digitāla pētījuma  
priekšmets var būt

objekts

attēls

skaņa

video

telpa, vide

teksts

metadati

## **Tekstizrāce - sakarību atrašana lielos teksta masīvos**

(«apslēptas» sakarības, «neparedzētas» attiecības)

### **Jēdzieni ar līdzīgu nozīmi:**

teksta analīze (*text analytics*)

makroanalīze (*macroanalysis*)

tāllasīšana (*distant reading*)

### **Pielietojuma jomas, piemēram:**

noskaņojuma analīze

tematu modelēšana

stilometrija

Kāpēc vajadzīga «mašīnlasīšana»...

Lielākā daļa grāmatu izzudīs uz visiem laikiem – un, sakot «lielākā daļa», mēs patiesībā izsakāmies neprecīzi: ja pieņemam, ka šodien 19. gs. angļu romānu kanonā ir apmēram divsimt nosaukumu, tad tie joprojām ir tikai 0,5 procenti no visiem publicētajiem romāniem.

Moretti, F. «The Slaughterhouse of Literature» (2000)



## PERIODIKA

Manas grāmatzīmes | Mani ieteiktie materiāli | Manas kolekcijas | Mani labojumi | Mani komentāri | Mani novērtētie | Mani saraksti



Izvērstā  
meklēšana

## IEROBEŽOT REZULTĀTUS

< Lapa 1 no 31108 > Atrasti: 622156 rezultāti Kārtot pēc: atbilstība Rādīt lapā: 20 40

## Pieejamība

- Bāreņdarbs 334296
- Pieejams bibliotē. 224596
- Publiski pieejams 63264

## Valoda

- Latviešu 614749
- Vācu 6977
- Spāņu 301
- Angļu 113
- Franču 5
- Krievu 5
- Igaņu 2
- Zviedru 2
- Čehu 1
- Latīņu 1

## Autors

- Dzelme, R. 343
- Pavasara, L. 331
- Zīle, D. 325
- A. L. 295
- Bāriņa, A. 290



## PAVASARIS

1940.04.20 Jaunākās Ziņas

gada laiki ir kā mazi pieturas punkti Mēs apstājami, ievēlam dziļāk dvašu un sakām — lūk, atkal pavasaris! Pavasaris ir pats ideālākais, maigākais, bet arī spēcīgākais gada laiks. Ziema ir tikai pratusi labāk vai sliktāk paglabāt tās vērtības, ko rudens tai nodevis. Rudens tikai savācis un briedējis



## PAVASARIS SĀCIES...

1967.03.23 Padomju Students

PAVASARIS SĀCIES... Э Хвоя рука в I моей руке. советы живущим в общежитии Хоть и весна на душе, возвращаясь поздно, не пой громко даже «Тишину» IDEĀLS STUDENTS Svetozora Zagorova zīm. SAKAS PAVASARA



## Latgale mana, tevi turos

1990.03.22 Vaduguns (Balvi)

ŠPERBERGA, Māriete

baltinaviešiem par 15910 darbu, svētku sajūtu un ierosmi' padomāt arī par citiem. . I Māriete ŠPERBERGA PAVASARIS



## PAVASARIS JAUNĀ IZDEVUMĀ

1987.05.29 Liesma (Valmiera)

PĒTERSONE, I.

PAVASARIS JAUNĀ IZDEVUMĀ lk .gadu pavasaris starp Себы, 'Rīāju 'un Burtnieku ezeru nāk ;kā jauna dabas ;grāmata citā izdevumā Nāļčpliešos zābaokos un aiz rokas ved vavari. я «Struaawmē» pavasĀar-is



## Rīgas Balss

1958.03.28 Rīgas Balss

fotostudija, L. Bites

L. Bites fotostudija PAVASARIS.

Atslēgvārda meklēšanas rezultāti tradicionālā digitālā kolekcijā

<https://www.sketchengine.eu/>

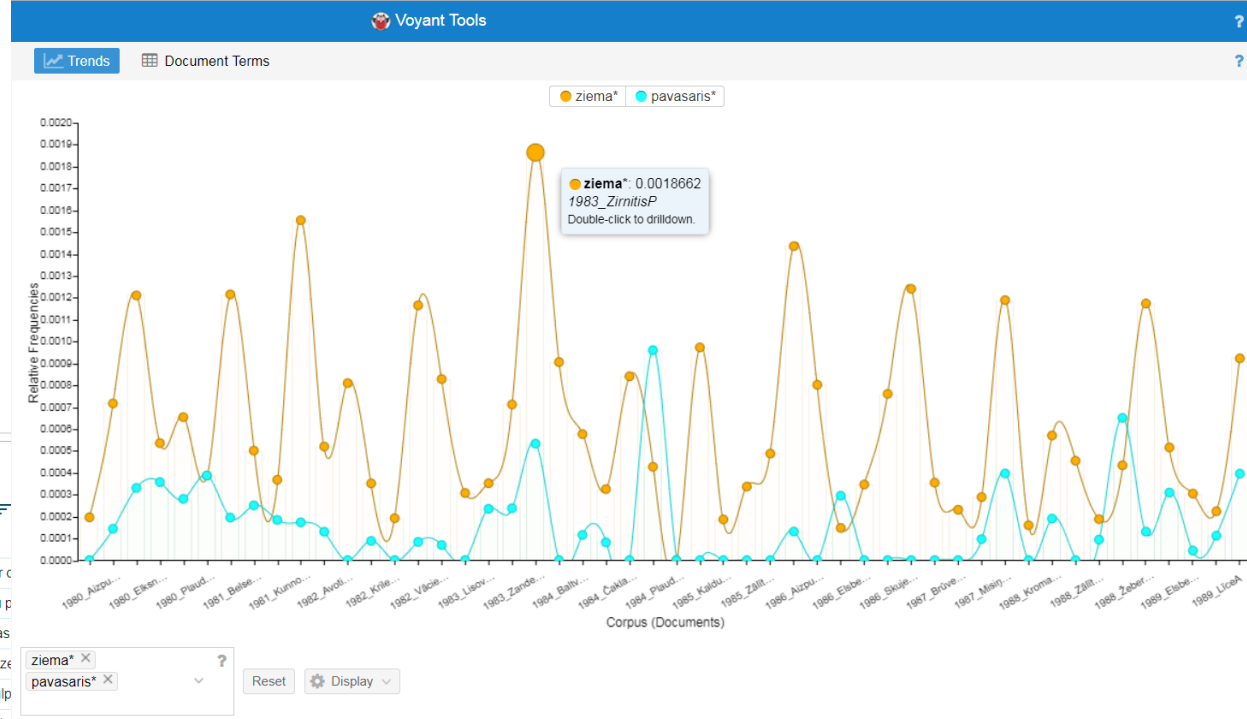
## CONCORDANCE

1980 gadu krājumi

lemma **pavasaris** 266 (480.19 per million)

Details	Left context	KWIC	Right context
1 doc#0 vējš nav nācis . Labi , ka atrnācis miers . Augumi gaiši un lokani Mūzikā sildās . Vai mēs no viena koka , No viena		<b>pavasara</b>	? Viena valoda mūsu rokām . 3. Puteņa stundā kļuva par c
2 doc#0 ūsus spārnus . Vējš , tikai vējš Tavas lūpas skāra . Skaistu Tev , rūgtais , ieraudzīt mani Pumpuru gaismā , Skaistu		<b>Pavasara</b>	mostošā kaislē . Tev , rūgtais rudens zēn , Novērst skatu p
3 doc#0 ļjavu nāk nāve – Basa , dzeltenā kleitā bez izkaptis . Tev ir skaista un godīga seja , nāvīt . Es tev ticu . Es ticu , ka		<b>pavasaros</b>	Mirst no mīlestības Un no ātruma – tāpat kā no mīlestības
4 doc#0 ūžas , Kamēr viņš pats Nāk smaidošai meitenei blakus pa pieneņu ļjavu Ar sauli kā dzīvību – Tieši . sejā . Viņš ir		<b>pavasari</b>	, Pavasari , kurš vakar beidzās . Puķu sieviņas pārdod dzē
5 doc#0 iēr viņš pats Nāk smaidošai meitenei blakus pa pieneņu ļjavu Ar sauli kā dzīvību – Tieši . sejā . Viņš ir pavasari ,		<b>Pavasari</b>	, kurš vakar beidzās . Puķu sieviņas pārdod dzeltenas tulp
6 doc#0 . Tātad es drīkstu aiziet ? Ne pirksts , ne pienešs kāts Nesalieksies – vilkt klāt Vai vismaz atgādināt Kopīgo		<b>pavasari</b>	. Laidies , baloniņ , laidies . . . Sažņaugti pirksti slaidie – Lā
7 doc#1 gām ; sit pie lo - ga zars . Laiks , izdalīts ar divi , tek lēnāk . Apnik brīdis , sa - pin sprīdis . Vēl pāris sekundes līdz		<b>pavasariem</b>	( stāv pulkstens , stāv ) . Laiks , izdalīts ar trīs , tek atpakaļ . Zars rūk par asnu , neatplaukst un satek sēklā . Un r
8 doc#1 milz , un abēlē ir tāds , kas piespiež novērst acis ( un paliek kluss – ko gaidi , valoda ? ) 1974 ČETRI GADALAIKI		<b>Pavasaris</b>	Kliedz dievam slava cīrulis kliedz aieļujā gaisā augstu Sit durvīs sirds kliedz sirds un sit kliedz durvīs sit ar savu
9 doc#1 ani rati Vijole un smiekli gausi Plūst un meklē manu ausi Ķekatnieki brauc pa ceļu Es ne acu nepaceļu 1970 Tā ir		<b>pavasara</b>	nakts tādej zemei tik silti . Zvaigzne staigā pa debesīm smarža pa zemes virsu . Cilvēki guļ un kas mīl pat tie jau
10 doc#2 c kādas intervijas . – Lai atceras , ka nevar savus zarus Bez Zēlastības tikai lauzt un lauzt . Ir jāļauj reizēm atnākt		<b>pavasariem</b>	Un jaunus pumpurus no jauna aust . Lai atceras , ka nevar savu aku Bez sāta katru dienu smelt un smelt , Ir jāļa
11 doc#2 ir uzdzīdis no sēklu spietiem ? Cik vēl liels tas nezināmais ikss , Gredzenus kas serdei apkārt liks ? SNIEGAINĀ		<b>PAVASARA</b>	DIENĀ Tik tuvu bij pienācis pavasars , Ka šķīta – ar roku var aizsniegt Tās bārkstis , ar kurām lazdu zars Aprīļa
12 doc#2 ķe , No ziedu šķīvjiem tauriņš ēd Un man uz mājām vajag kuģot – Es nedrīkstu to nokavēt . VIŽŅU ZVĀRGULIŠI		<b>Pavasaris</b>	nāk , bet manī Vēl arvienu rudens sēž , Vižņu zvārgulišiem zvana , Nost no kļāvām lapas plēš Un ar savu roku m
13 doc#2 izīrksti dzēš ... Tumsa , tumsa , migla , migla Prom no zariem ziedus mēž . KAUT REIZI VĒL Šo izstādi , ko atver		<b>pavasaris</b>	, Kaut reizi vēl bez steigas izstaigāt ! Kā toreiz bērniņā , kad dienas bij tik garas , Kad vakars neskreja tik strauji
14 doc#2 , lidot , kustēties Un saldā aizmirstībā sviļot , un dziedāt , Un badīties , un ļīpu gaisā sviest . Šo izstādi , ko atklāj		<b>pavasaris</b>	, Bez steigas izstaigāt kaut reizi vēl ! Lai tad , kad jāceļas būs Lētei pāri , Nav neredzētā žēl . IT KĀ NO JAUNĪBA
15 doc#2 . Un viņi bij jauni un laimīgi – Kā zeme , kā ūdens , kā zāle . Un visi liegumi piepeši Bij kļuvuši vārgi un bāli . Pret		<b>pavasari</b>	– es sapratu – Man veltīgi cīnīties būtu . Ar padomiem saviem pat zvīrbuļiem , Pat knišļiem es smiekliņa kļūtu . A
16 doc#2 r . Stāv gar mūsu tīrumiem un sētām , Siltu ēnu ceļiem pāri met , Dziedinošu mieru pārklāj rētām , Pretim jaunam		<b>pavasariem</b>	ved . Skuju raksti mātes dotā cimdā , Jānis zaļām ozolapām vīts , Tāšu svārkos cauri gadusiemtiem Tautas dzies
17 doc#3 na kā puķe . Viņa stāvēja viena , baroja pienešus suķim . Suķis sukāja iekšā visu , ko priekšā tam cēla . Bija vēls		<b>pavasaris</b>	, zāle kā traka zēla . Taisījās ļaudis ļlaut sienu , lai kļūtu par barību puķes . Skuķe baroja suķi un jutās briesmīgi
18 doc#3 pēc nenositu to smurguli , kas spļāva man . Lai dzīvo viņš . Un lai tev arī liek kādu nieku padomāt . Jo kā lai notic		<b>pavasariem</b>	, ja pavasaris vēl nav klāt ? Cilvēks nāk un nodrāž mietu . Cilvēks iet un mietu sprauž . Kas gan tā par burvju viet
19 doc#3 smurguli , kas spļāva man . Lai dzīvo viņš . Un lai tev arī liek kādu nieku padomāt . Jo kā lai notic pavasariem , ja		<b>pavasaris</b>	vēl nav klāt ? Cilvēks nāk un nodrāž mietu . Cilvēks iet un mietu sprauž . Kas gan tā par burvju vietu , kur tos mie
20 doc#4 Gaišzila jūra un debesu lauks ... Vēl tavas trīsošās rokas nezina , Kādu glābēju palīgā saukt . Kas var izglābt no		<b>pavasara</b>	Un no skaudras vanaga acs ? Kam ir dota tik savāda vara – Dedzināt tā , lai nesakarsi ? Tavas acis ir rudens jūr

Rows per page: 20 1–20 of 266



<https://voyant-tools.org/>

Atslēgvārda meklēšanas rezultāti, pētot atsevišķu datu kopu ārpus digitālās kolekcijas

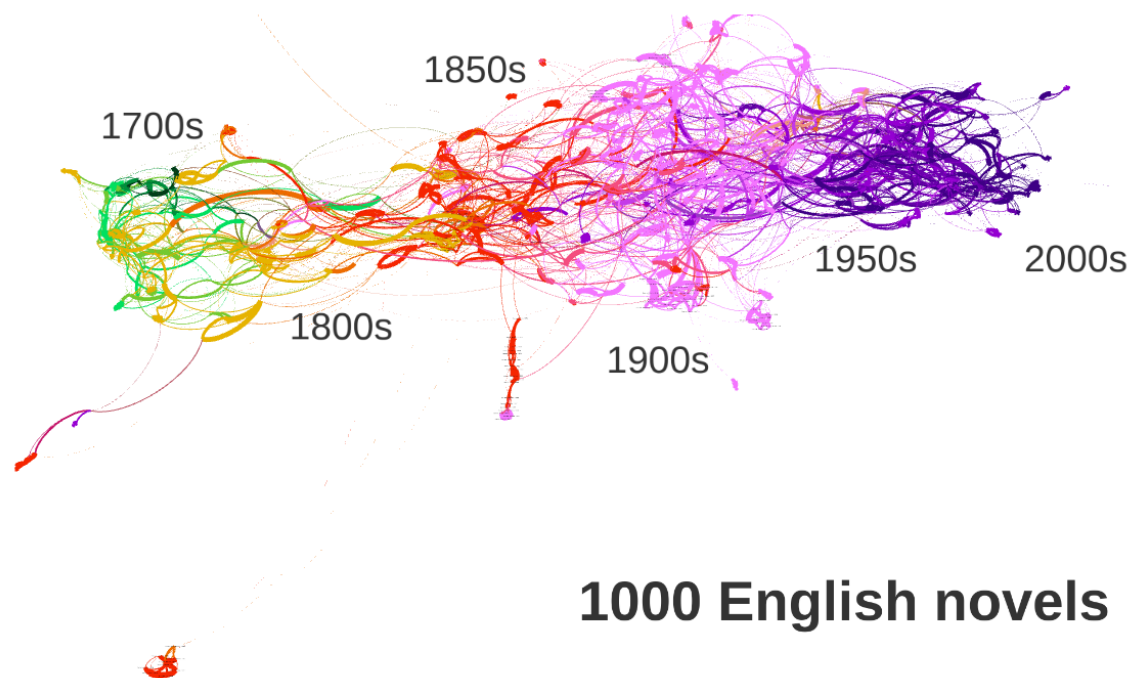
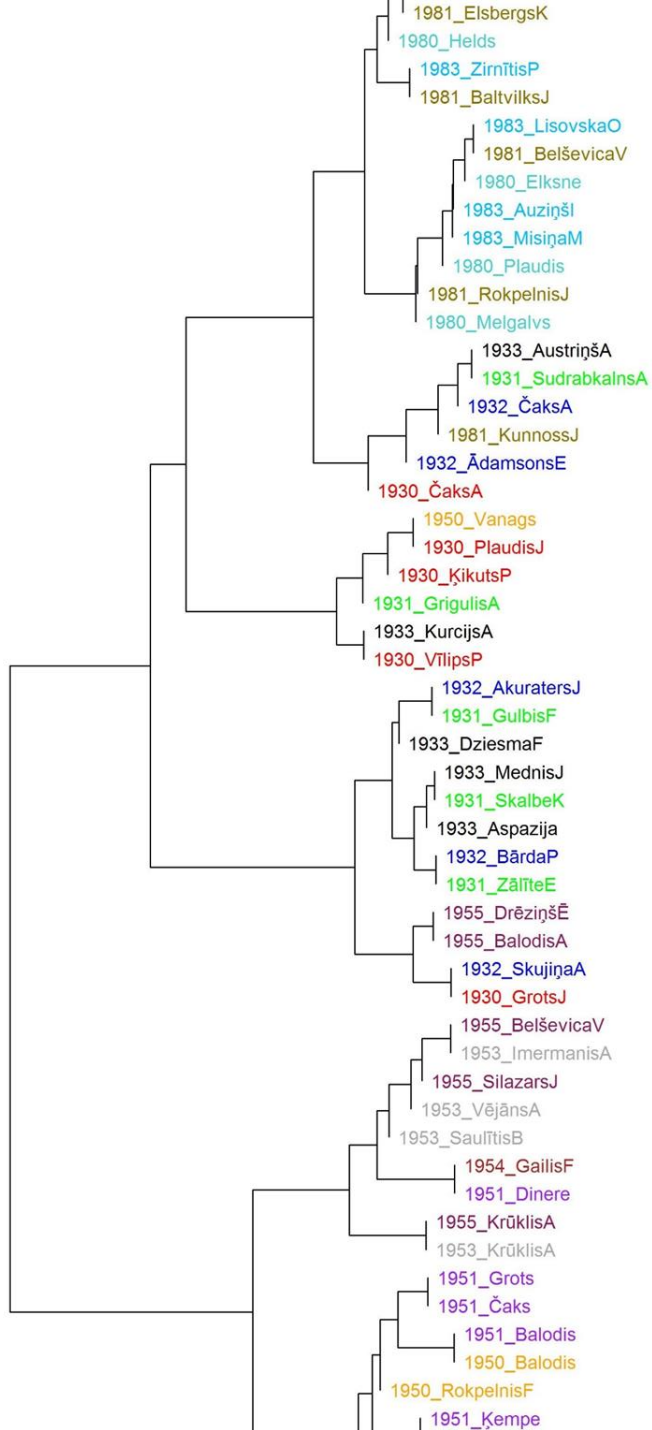
ASV prezidenta D. Trampa Twitter ierakstu analīze

<https://medium.com/swlh/analyzing-trumps-tweets-5368528d2c90>



# Stilometrija (*stylometry*)

Tekstizraces un mašīnmācīšanās algoritmu lietojums filoloģijā, literatūras pētniecībā, statistiska metode, kas analizē vienlaikus liela vārdu (vai citu tekstvienību) skaita uzvedību, meklējot sakarības, tendences.



1000 English novels

# Ko var pētīt ar tekstizraces palīdzību (daži piemēri)

- Valodas attīstība laika gaitā, dažādās demogrāfiskās grupās
- Konkrētu tematu aktualizēšanās un noriets laika gaitā
- Sabiedrības attieksme pret dažādām parādībām, t.sk. laika gaitā un salīdzinot dažādas grupas
- Savstarpējas ietekmes tautu un autoru vidū
- Vai ģeniāli literāri darbi rada skolas un tradīcijas
- Kāda tieši ir atšķirība starp kanonisku un populāru literatūru



Pastāv uzskats, ka bibliotēkās, arhīvos, muzejos nav vērts vai nav iespējams ieviest rīkus padziļinātai digitālai pētniecībai, jo katram pētniekam vajadzīgs kaut kas cits.

Šajā viedoklī ir sava daļa taisnības. Tomēr, tāpat kā mums noder sabiedriskais transports, lai arī mēs katrs dzīvojam citā dzīves vietā, ir atrodamī tādī digitālās pētniecības rīki un risinājumi, kas ir pietiekami universāli, elastīgi un var nederēt daudziem, pat tad ja neaizved «līdz pašam dzīvokļa sliekšnim».

Piemēram:

- iespēja attēlot meklētā jēdziena lietojuma dinamiku uz laika skalas (*n-gram viewer*);
- iespēja no digitālās bibliotēkas izgūt objektu kopas, lai tās padziļināti pētītu (API piekļuve vai iespēja lejupielādēt *txt*, *xml* failus vai tādus failus, kurus var pārveidot šādos formātos);
- iespēja monogrāfiju vai periodikas digitālajās kolekcijās automātiski atlasīt specifiskus segmentus, piem., attēlus, reklāmas laukus, lai ar tiem varētu strādāt uz viena, pārskatāma darba galda.



Iesaisties digitālās pētniecības tīklā!  
digitalhumanities.lv



# Starptautiska konference

“Kultūras mantojums digitālā vidē: veidojot sadarbības tīklu Baltijas reģionā”

2019. g. 21.-22. maijā Latvijas Nacionālajā bibliotēkā

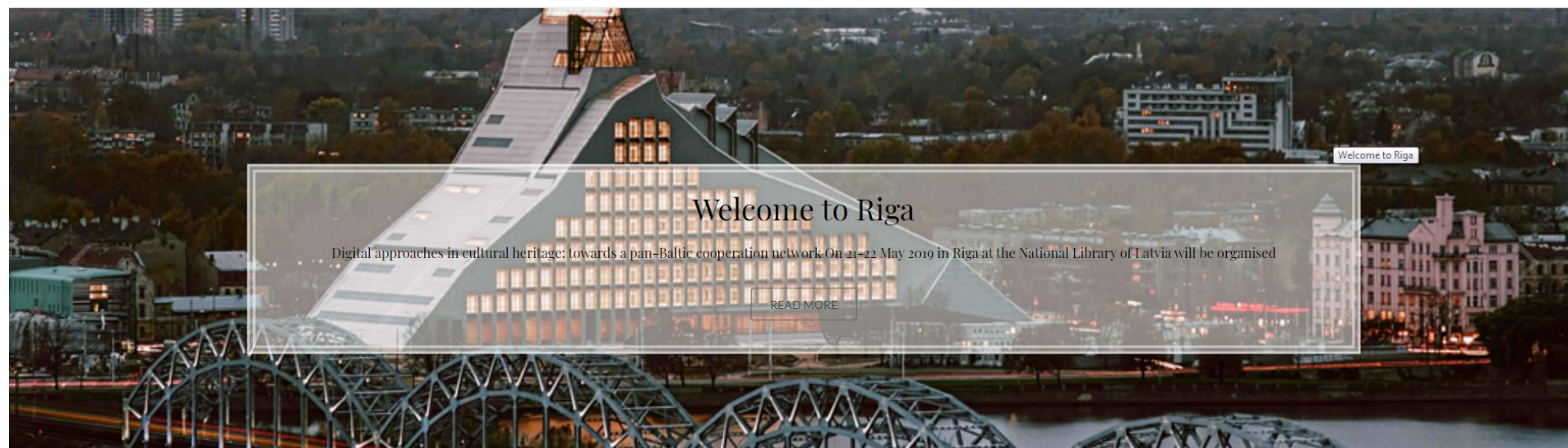
Pieteikšanās: <https://dach2019.lnb.lv/>

International conference  
Digital approaches in cultural heritage:  
towards a pan-Baltic cooperation network

21–22 May 2019, Riga, Latvia



HOME REGISTRATION PROGRAMME PARTNERS RIGA CONTACTS



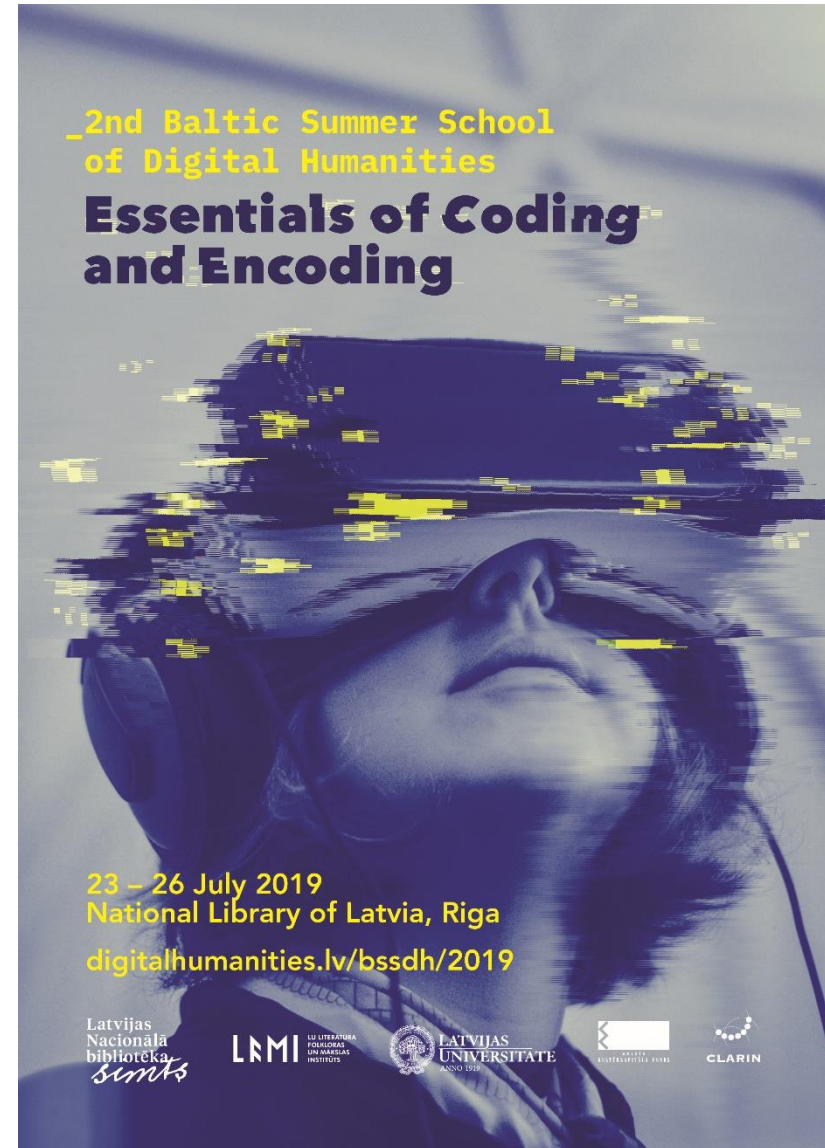
# Otrā Baltijas Digitālo humanitāro zinātņu vasaras skola

2019. g. 23.-26. jūlijā

Latvijas Nacionālajā bibliotēkā

Pieteikšanās:

<http://www.digitalhumanities.lv/bssdh/2019/>



Ikgadējā *Digital Humanities in Nordic Countries* konference DHN2020

2020. g. 17.–19. martā

Latvijas Nacionālajā bibliotēkā

Pieteikšanās vēl nav atklāta

Referātu iesniegšana no 2019. g. augusta

DHN



Paldies!