CLARIN

Vienota valodas resursu un tehnoloģiju infrastruktūra

Inguna Skadiņa
CLARIN-LV nacionālā koordinatore

Latvijas Universitātes
Matemātikas un informātikas institūts

Zinātnieku brokastis
RSU
2023. gada 6. decembrī

CLARIN

Vienota valodas resursu un tehnoloģiju infrastruktūra

CLARIN (Common Language Resources and Technology Infrastructure) ir digitāla **Eiropas pētniecības infrastruktūra** (ERIC)

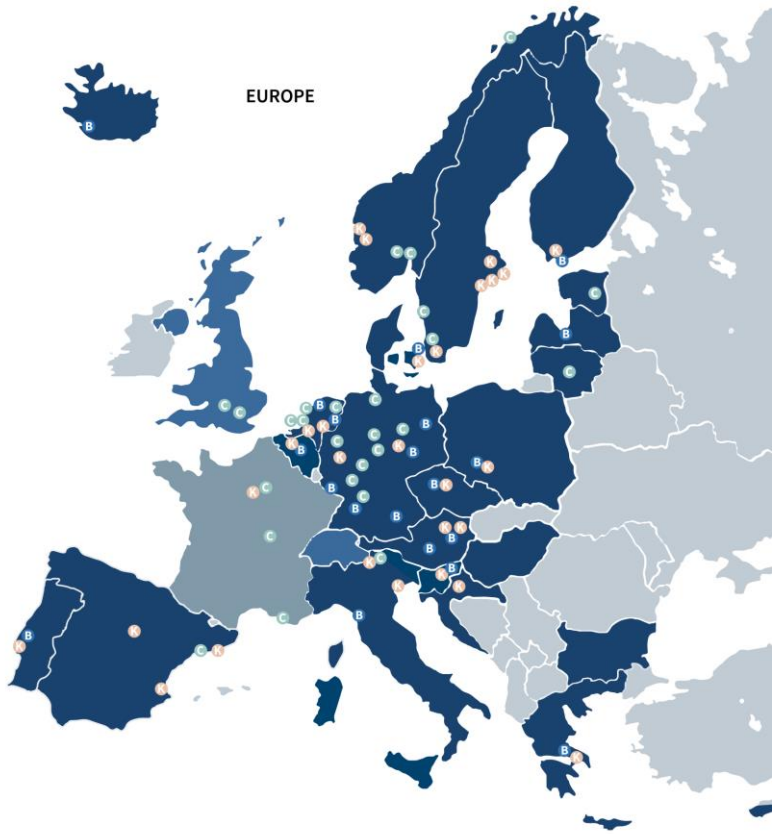
Tās pamatā ir ideja, ka

visi **digitālie valodas resursi un rīki** Eiropā un citur pasaulē ir pieejami

visu jomu, it īpaši **humanitāro un sociālo zinātņu** zinātniekiem un studentiem, izmantojot **vienotu pierakstīšanos.**

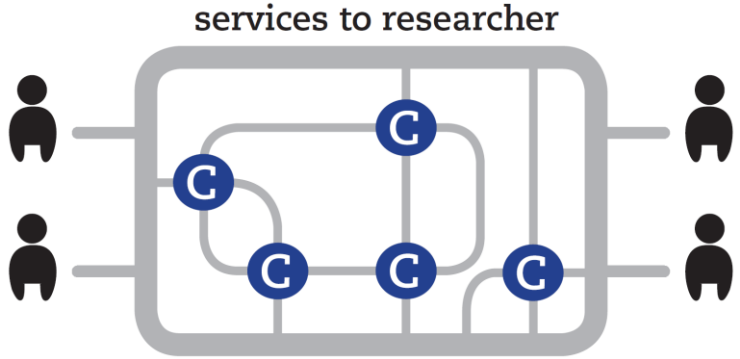
CLARIN

CLARIN – dalīta Eiropas pētniecības infrastruktūra



- 23 + 3 valstis
- > 60 reģistrētu centru
- Zināšanu centru tīkls (K-centri)
- Vienkārša un **ilgtspējīga** piekļuve **digitālajiem valodas datiem**
- Mūsdienīgi **valodas apstrādes rīki**, valodas datu izpētītei un analīzei neatkarīgi no to atrašanās vietas
- Cieša sasaiste ar Atvērtās zinātnes programmu, fokuss uz **sadarbspēju**

FAIR principles for open language resources



- **F** – *findability*:
  existence at a data **provider**, discovery, **persistence**, long-term **preservation**
- **A** – *accessibility*:
  open data / access
- **I** – *interoperability*:
  common **terminology**, data **models** and **formats** (for both humans & machines)
- **R** – *reusability*:
  content, **documentation**, provenance, **versioning**, licencing

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018, 2016

CLARIN

Par CLARIN

CLARIN (Common Language Resources and Technology Infrastructure –Vienota valodas resursu un tehnoloģiju infrastruktūra) ir dalīta Eiropas pētniecības infrastruktūra (ERIC), kurā integrētas zināšanas, valodas resursi un rīki, repozitoriji un servisi, ko piedāvā daudzie CLARIN centri un nacionālie konsorciji.

CLARIN mērķis ir novērst sadrumstalotību valodas resursu jomā un padarīt valodas resursus pieejamus visu jomu, it īpaši humanitāro un sociālo zinātņu zinātniekiem un studentiem. Lai to īstenotu, izveidota kopīga izkliedēta infrastruktūra, kurā integrēti valodas resursi un rīki no visas Eiropas. CLARIN piedāvā ilgtermiņa risinājumus un servisus digitālo valodas datu un rīku izvietošanai, savienošanai un analīzei.

Latvija CLARIN iniciatīvā piedalās kopš tās pirmsākumiem. 2008. gadā LU MII kļūst par CLARIN sagatavošanas posma projekta konsorcija dalībnieku, būtisku finansiālu atbalstu dalībai projektā sniedz Izglītības un zinātnes ministrija. 2016. gada 1. jūnijā Latvija pievienojas CLARIN ERIC. Izglītības un zinātnes ministrija uzticējusi LU Matemātikas un informātikas institūtam pārstāvēt Latviju CLARIN ERIC un vadīt Latvijas CLARIN konsorciju. 2019. gadā CLARIN-LV pievienojas CLARIN zināšanu centram morfoloģiski bagātām valodām SAFMORIL. 2020. gada martā tiek izveidots CLARIN-LV repozitorijs, kurā tiek reģistrēti latviešu valodas resursi un rīki. Esot CLARIN ERIC konsorcijā, CLARIN-LV ir apņēmusies ilgstoši rūpēties un saglabāt repozitorijā deponētos digitālos valodas resursus un rīkus. CLARIN-LV īsteno CLARIN ERIC misiju Latvijā, vācot, dokumentējot, kurējot un nodrošinot ilgtermiņa piekļuvi digitālajiem latviešu valodas resursiem un rīkiem. 2022. gada janvārī, konsorcijā apvienojoties septiņiem partneriem, tiek noslēgts CLARIN nacionālā konsorcija līgums, kurā CLARIN-LV konsorcijs vienojas par CLARIN-LV darbību un turpmāku attīstību. 2023.gada janvāri CLARIN-LV repozitorijs iegūst B-centra statusu un saņem CoreTrustSeal sertifikātu kā uzticams datu repozitorijs.

CLARIN-LV regulāri rīko konferences un seminārus, kuros iepazīstina ar dažādiem valodas resursiem un rīkiem un to lietojumu pētniecībā un valodu tehnoloģiju izveidē. Latvijā paveiktais apkopots vairākās publikācijās.

Kopš 2018. gada CLARIN Latvija darbību atbalsta ERAF projekts "Latvijas Universitāte un institūti Eiropas pētniecības telpā - ekselence, aktivitāte, mobilitāte, kapacitāte", kā arī projekti "Humanitāro zinātņu digitālie resursi: integrācija un attīstība" (5.10.2020-5.10.2022) , "Atvērtas un FAIR principiem atbilstīgas digitālo humanitāro zinātņu ekosistēmas attīstība Latvijā" (kopš 2022. gada decembra), "Mūsdienu latviešu valodas izpēte un valodas tehnoloģiju attīstība" (kopš 2022. gada) un "Valodu tehnoloģiju iniciatīva" (kopš 2023. gada).

CLARIN-LV piedalās ikgadēja CLARIN konferencē

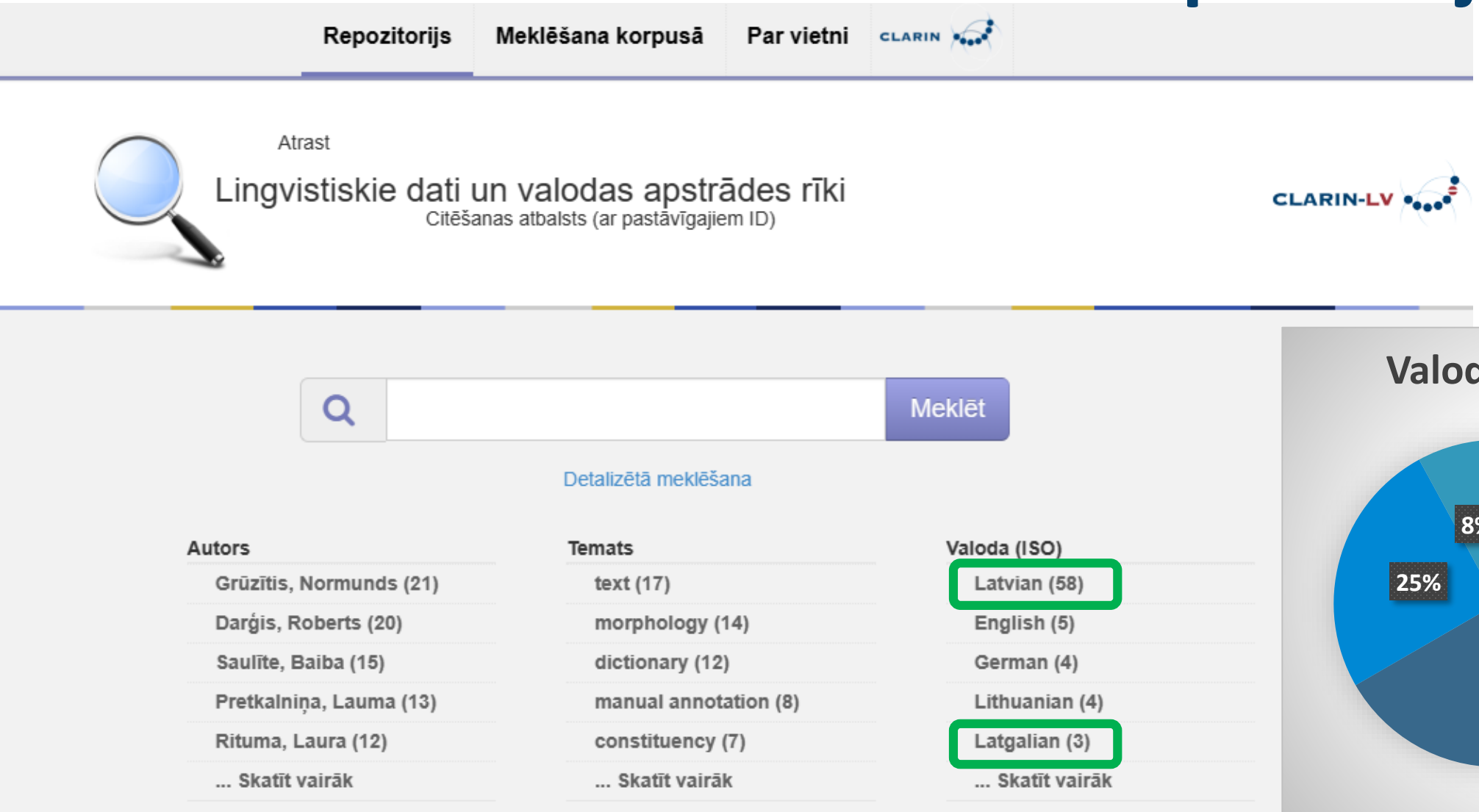
No 16. oktobra līdz 19. oktobrim Lēvenē notika gadskārtējā Eiropas pētniecības infrastrutūras CLARIN konference.

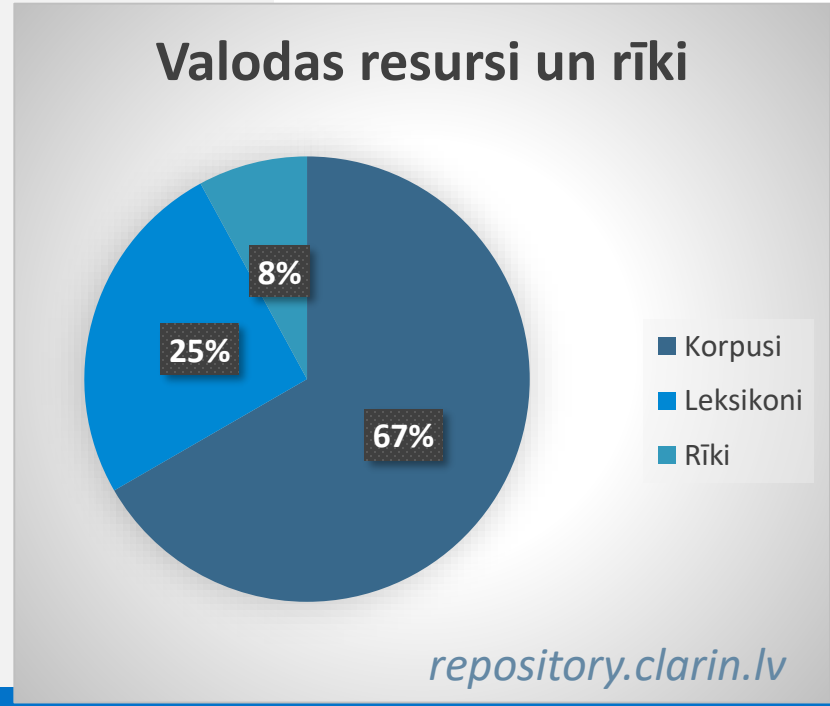
CLARIN B-centre

CLARIN K CENTRE

www.clarin.lv
info@clarin.lv

CLARIN-LV valodas resursu repozitorijs





CLARIN-LV valodas resursu repozitorijs

Kas jauns

LexicalConceptualResource — CLARIN Centre Of Latvian Language Resources And Tools

Dictionary of Contemporary Latvian Language (MLVV) (2023-09-21)

Autors (-i):

Jērāne, Santa ; et al.

▶ show everyone

Apraksts:

"Contemporary dictionary of Latvian language" (MLVV), which is developed by the UL Latvian Language institute, is a new explanatory dictionary based on Latvian language materials obtained during the last decade. The analysis ...

Šajā vienumā ir 1 fails (7.05 MB).

Publicly Available

Corpus — CLARIN Centre Of Latvian Language Resources And Tools

Corpus of Latvian PhD Theses (Disertācijas)

Autors (-i):

Darģis, Roberts

Apraksts:

The corpus consists of PhD theses and abstracts published in the University of Latvia, Riga Technical University, Riga Stradins University and Liepaja University until 2020.

Šajā vienumā ir 3 faili (127.73 MB).

Publicly Available

CLARIN-LV

repository.clarin.lv

CLARIN-LV valodas resursu repozitorijs

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

LVMED: Latvian Speech Transcripts of the Medical Domain

(AiLab IMCS UL / 2021-09-30)

Autors (-i):

Auziņa, Ilze ; et al.

▶ show everyone

Šajā vienumā ir 1 fails (267.18 KB).

Publicly Available

SINGLE-WORDS ✓ MULTI-WORDS ✓

reference corpus: Saeima

	Word		Word		Word		Word		Word
1	plauša	11	CT	21	patoloģisks	31	pleira	41	perēklis
2	limfmezgls	12	urīnpūslis	22	žultspūslis	32	dobums	42	iekava
3	akna	13	virsniere	23	daiva	33	mililitrs	43	sieniņa
4	centimetrs	14	liesa	24	gluds	34	kontrastēt	44	aksiāls
5	homogēns	15	dziedzeris	25	izmeklējums	35	ieslēgums	45	žultsvads

CLARIN-LV

CLARIN-LV valodas resursu repozitorijs

Corpus

CLARIN Centre Of Latvian Language Resources And Tools

LVMED: Latvian Speech Transcripts of the Medical Domain

(AiLab IMCS UL / 2021-09-30)

Autors (-i):

Auziņa, Ilze ; et al.

▶ show everyone

Šajā vienumā ir 1 fails (267.18 KB).

Publicly Available

					vecuma normas robežās. </s><s> labajā pusē vecs l
					āta pa kreisi, truncus pulmonalis trīs komats trīs cent
23	doc#89	svaigu infiltrāciju neredz . </s><s> pneimofibroze. </s><s>	sirds	palielināta pa kreisi, aortа vecuma robežā . </s><s> trahejā
24	doc#91	sklerozētu sienu. </s><s> un tad secinājumā lūdzu rakstām	sirds	neizvērtējama punkts </s><s> un tad apr* secinājumā lūdzu
25	doc#94	s> pastiprināts plaušu zīmējums piesakņu rajonā . </s><s>	sirds	, aorta vecuma normā . </s><s> sinusi brīvi. </s><s> vēdera
26	doc#103	><s> nākamā rindā lūdzu rakstām nākamajā rindā </s><s>	sirds	paplašināta šķērsizmērā , vairāk pa kreisi. </s><s> vēl atzīn
27	doc#118	·<s> aorta nav paplašināta. </s><s> nākamā rindā </s><s>	sirds	nav paplašināta šķērsizmērā . </s><s> un tas arī viss paldie
28	doc#134	><s> saknes sabiezētas , salīdzinoši kreisā sakne. </s><s>	sirds	palielināta pa kreisi, aortа vecuma robežā . </s><s> kontrol
29	doc#138	hiperplastisks etmoidīts un punkts </s><s> plaušu lauki un	sirds	bez redzamas novirzes no normas. </s><s> paraksts </s><
30	doc#141	·rencējas punkts šķidrumu pleiras telpās nekonstatē punkts	sirds	nav paplašināta šķērsizmērā . </s><s> aorta nav paplašināt
31	doc#145	s> plašu emfizēma. </s><s> difūza pneimofibroze. </s><s>	sirds	nav palielināta. </s><s> aorta plata, sklerozēta. </s><s> vir
32	doc#154	·glu kalcinācija kreisā pusē. </s><s> tālāk rakstām </s><s>	sirds	šķērsizmērā nav paplašināta. </s><s> aorta sklerotizēta. </

CLARIN-LV

CLARIN-LV valodas resursu repozitorijs

Tool Service

CLARIN Centre Of Latvian Language Resources And Tools

RUTA:MED – Dual Workflow Medical Speech Transcription Pipeline and Editor

(AiLab IMCS UL / 2022)

Author(s):

Znotiņš, Artūrs ; Darģis, Roberts ; Grūzītis, Normunds ; Bārzdiņš, Guntis and Goško, Didzis

This item contains no files.

Saglabāt | 5s | B | I | U | x^2 | x_2 | ☰

MR prostatai ar k/v.

Prostata 3,9 x 3,3 x 3,8 cm, tilpums 26,15 cm^3.

PSA līmenis nav zināms.

Dziedzera pārejas daļā redzami labdabīgas hiperplāzijas mezgli.

Urīnpūslis gludām sieniņām, bez intralūmenāliem ieslēgumiem.

Patoloģiska lieluma l/m nesaskatu.

Slēdziens

Labdabīga prostatas hiperplāzija.

Pašreiz nav datu par augstas malignitātes tumora audiem prostatā.

0.00s 4.72s 16.41s 36.58s 68.44s 72.16s 75.92s 77.74s 86.06s

▶ 00:00:06 / 00:01:32

CLARIN-LV

CLARIN-LV valodas resursu repozitorijs

Kas jauns

LexicalConceptualResource — CLARIN Centre Of Latvian Language Resources And Tools

Dictionary of Contemporary Latvian Language (MLVV) (2023-09-21)

Autors (-i):

Jērāne, Santa ; et al.

▶ show everyone

Apraksts:

"Contemporary dictionary of Latvian language" (MLVV), which is developed by the UL Latvian Language institute, is a new explanatory dictionary based on Latvian language materials obtained during the last decade. The analysis ...

Šajā vienumā ir 1 fails (7.05 MB).

Publicly Available

Corpus — CLARIN Centre Of Latvian Language Resources And Tools

Corpus of Latvian PhD Theses (Disertācijas)

Autors (-i):

Darģis, Roberts

Apraksts:

The corpus consists of PhD theses and abstracts published in the University of Latvia, Riga Technical University, Riga Stradins University and Liepaja University until 2020.

Šajā vienumā ir 3 faili (127.73 MB).

Publicly Available

LexicalConceptualResource

CLARIN Centre Of Latvian Language Resources And Tools

Dictionary of Contemporary Latvian Language (MLVV) (2023-09-21)

Author(s):

Jērāne, Santa ; et al.

▶ show everyone

Description:

"Contemporary dictionary of Latvian language" (MLVV), which is developed by the UL Latvian Language institute, is a new explanatory dictionary based on Latvian language materials obtained during the last decade. The analysis ...

This item contains 1 file (7.05 MB).

Publicly Available

This item is Publicly Available and licensed under:
Creative Commons - Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Name	mlvv_2023_4_tei.zip
Size	7.05 MB
Format	application/zip
Description	MLVV open data in the TEI/XML format (https://tei-c.org/release/doc/tei-p5-doc/en/html/DI.html)
MD5	cef0a6b5399382c98b294b4406c2d77e

Download file Preview

Licences un to pārvaldīšana

ToolService

CLARIN Centre Of Latvian Language Resources And Tools

Ilvars - Latvian Male VITS Text-to-Speech Model (vers. 2023)

Autors (-i):

Darģis, Roberts and Auziņa, Ilze

Apraksts:

A neural model for text-to-speech (TTS) synthesis in Latvian. Trained using VITS on a 25-hour speech corpus of audiobooks read in a male voice. Available for academic and non-commercial purposes via an API. To get access ...

Šajā vienumā ir 1 fails (376.17 MB).

Restricted Use

Licences un to pārvaldīšana

ToolService

Ilvars - Latvian Male VITS Text-to-Speech Model (vers. 2023)

Autors (-i):

Darģis, Roberts and Auziņa, Ilze

Apraksts:

A neural model for text-to-speech (TTS) synthesis in Latvian. Trained using audiobooks read in a male voice. Available for academic and non-commerc

Šajā vienumā ir 1 fails (376.17 MB).

Licenču pārvaldīšana

All Licenses | Define License | Define License Label

License Name: CLARIN RESTRICTED

Definition (URL): https://www.kielipankki.fi/lic/example-license-clarin-res-id-by-[

Confirmation: Ask always

Label: RES

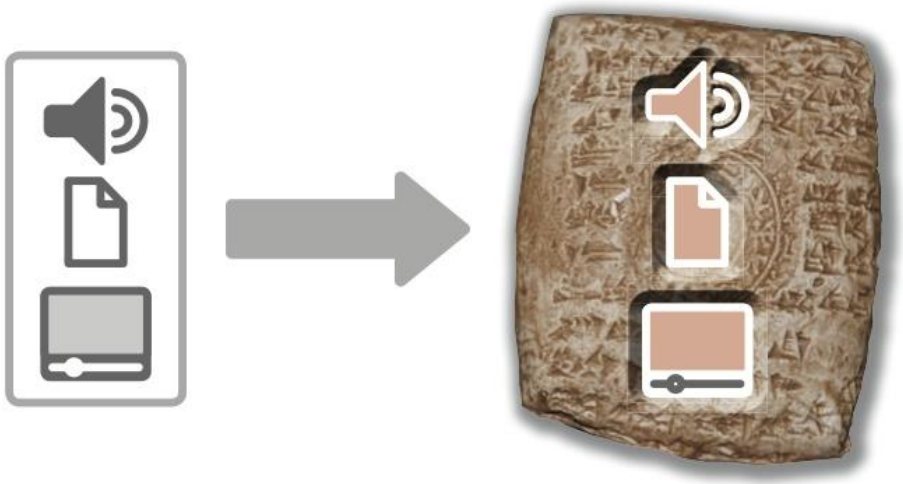
Extended Label(s): BY, SA, NC, ReD

Papildu nepieciešamā lietotāja informācija:

- [x] Lietotājs saņems e-pasta ziņojumu ar lejupielādes instrukcijām.
- [x] Lietotāja vārds
- [ ] Dzimšanas datums
- [x] Adrese
- [x] Valsts
- [ ] Vaicāt lietotājam citu e-pasta adresi
- [ ] Jautāt lietotājam par organizāciju (nav obligāti).
- [x] Jautāt lietotājam par organizāciju (obligāti).
- [x] Pajautājiet lietotājam par viņa nodomiem ar vienumu.

Datu uzglabāšanas servisi: deponēšana

- Viens no CLARIN pamatuzdevumiem ir nodrošināt, ka valodas resursi tiek arhivēti, tie ir **pieejami ilgterminā** un tie ir **uzticami**
- Tāpēc CLARIN centri piedāvā valodas resursu uzglabāšanas (deponēšanas) pakalpojumu



https://www.clarin.lv/attachments/CLARIN%202020_resursu_iesniegsana.pdf

Ilgtspēja

Core Trust Seal and B-centre certification

[Source: CoreTrustSeal Requirements 2020-22]

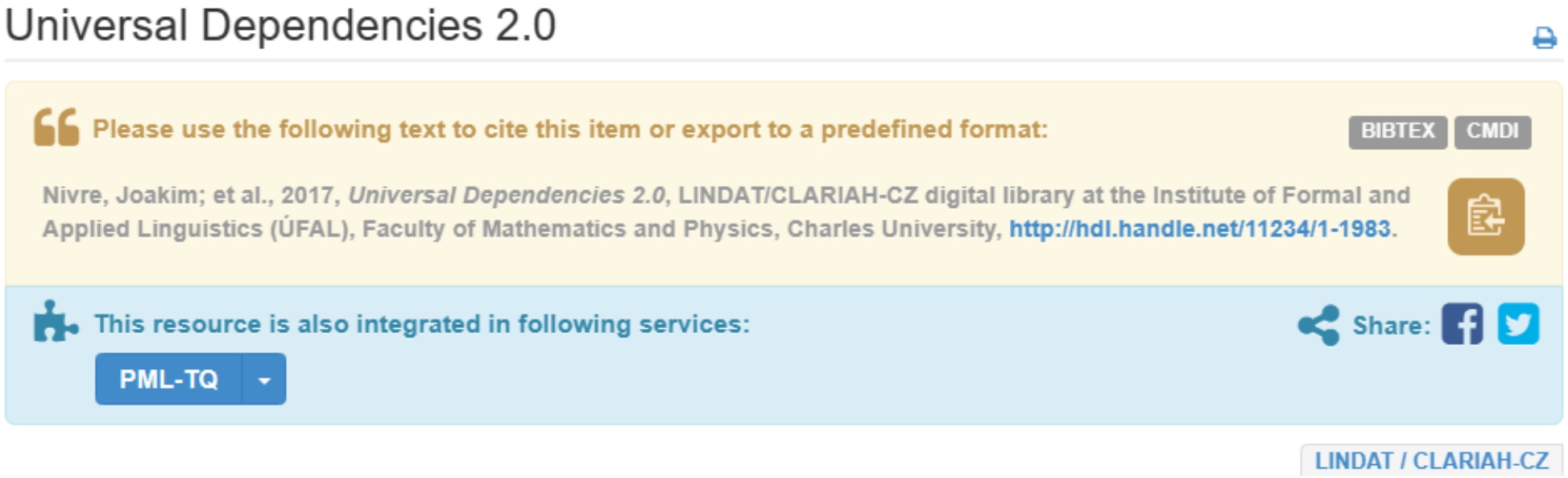
CoreTrustSeal self-assessment template

The CoreTrustSeal self-assessment includes 16 requirements (R1–R16) plus R0 providing contextual information on the repository. This template contains all requirements along with guidance for formulating self-assessment statements and gathering related evidence. Along with the self-assessment statements, the applicant must indicate a compliance level for R1–R16:

0. Not applicable
1. The repository has not considered this yet
2. The repository has a theoretical concept
3. The repository is in the implementation phase
4. The guideline has been fully implemented in the repository

The CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022 document contains more information on e.g. missing evidence, understandability, non-English language documentation, sensitive/internal documentation, structure and length of responses, and further guidance for certain requirements. Please read it before starting!

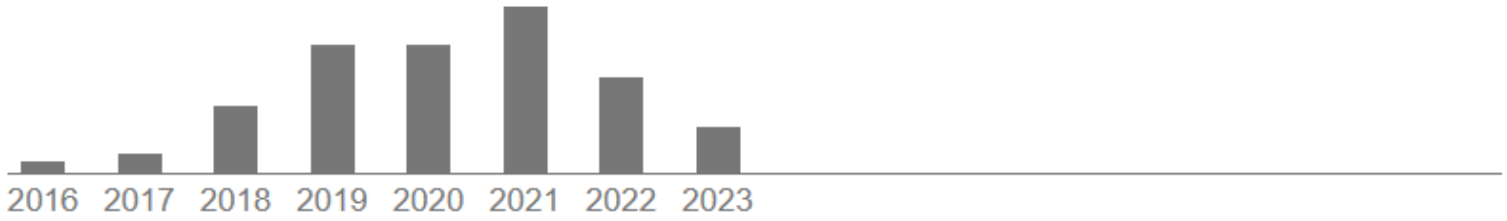
Citēšana un unikālie identifikatori



Citēšana

Universal Dependencies 2.0

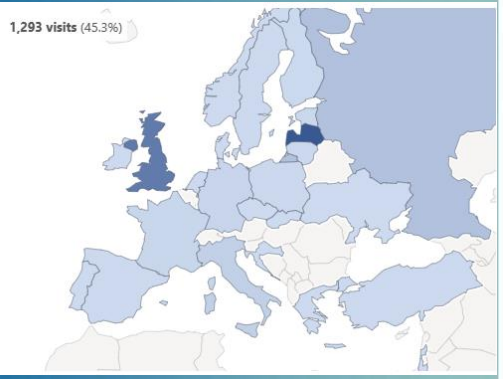
Autori	Joakim Nivre, et al.
Publicēšanas datums	2017/3
Avots	http://hdl.handle.net/11234/1-1983
Apraksts	UNIV-ST-ETIENNE| ENS-LYON| UNIV-LYON3| UNIV-PARIS7| ENS-PARIS| UNIV-TLN| PRES_CLERMONT| CNRS| INRIA| UNIV-AMU| UNIV-LYON2| UNIV-PARIS3| LATTICE| MODYCO| LLF| LPP| CERHAC| UNIV-LORRAINE| INRIA2| LORIA| LORIA-NLPKD| PSL| USPC| LIS-LAB| IHRIM| UNIV-PARIS-LUMIERES| UDL| UNIV-PARIS| UP-SOCIETES-HUMANITES| UNIV-PARIS-NANTERRE| IFRA-NIGERIA
Atsauču kopskaits	Minēts 312



CLARIN

Populārākie valodas resursi

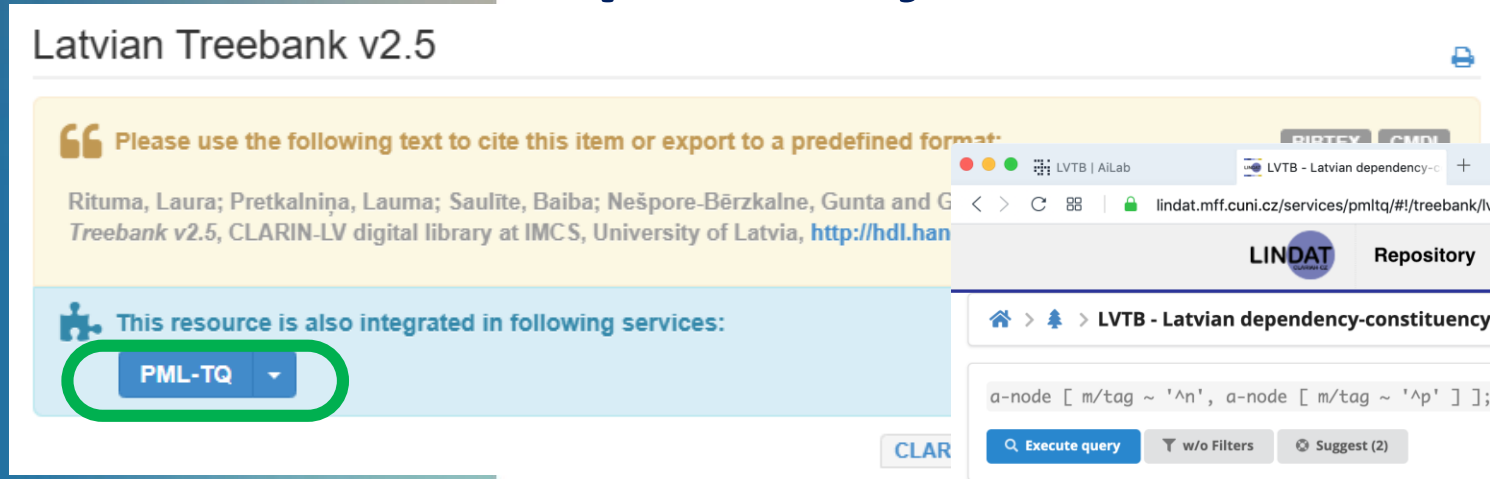
Visitor Map

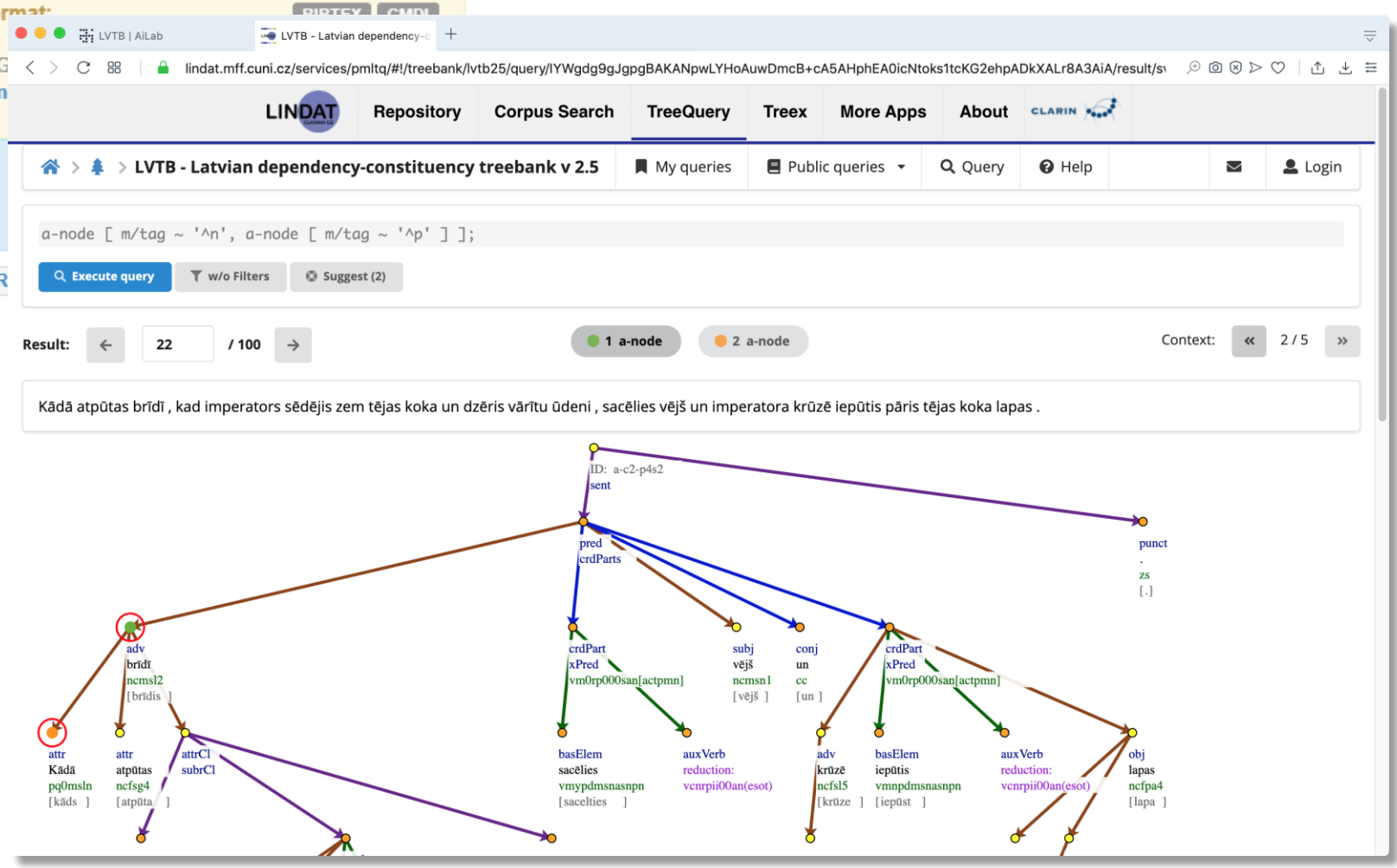


Item/Handle	Number of views
Tēzaurs.lv 2020 (Spektors, Andrejs et al) (20.500.12574/9)	5352 [+bots: 6275]
Latvian Treebank v2.5 (Rituma, Laura et al) (20.500.12574/10)	3447 [+bots: 4109]
Tēzaurs.lv 2022 (Spektors, Andrejs et al) (20.500.12574/66)	3296 [+bots: 3494]
Balanced Corpus of Modern Latvian (LVK2018) (Levāne-Petrova, Kristīne et al) (20.500.12574/11)	3093 [+bots: 3206]
LVBERT - Latvian BERT (Znotiņš, Artūrs) (20.500.12574/43)	2967 [+bots: 3574]
NLP-PIPE: Latvian NLP Tool Pipeline (Znotiņš, Artūrs et al) (20.500.12574/4)	2685 [+bots: 2857]
The Corpus of Early Written Latvian (Andronova, Everita et al) (20.500.12574/12)	2514 [+bots: 2882]
Rainis (Spektors, Andrejs et al) (20.500.12574/41)	2413 [+bots: 2820]
Corpus of Latvian Pandemic Diaries 2020–2021 (Reinsone, Sanita et al) (20.500.12574/48)	2403 [+bots: 2800]
Latvian AMR Sembank (Znotiņš, Artūrs et al) (20.500.12574/40)	2360 [+bots: 2705]

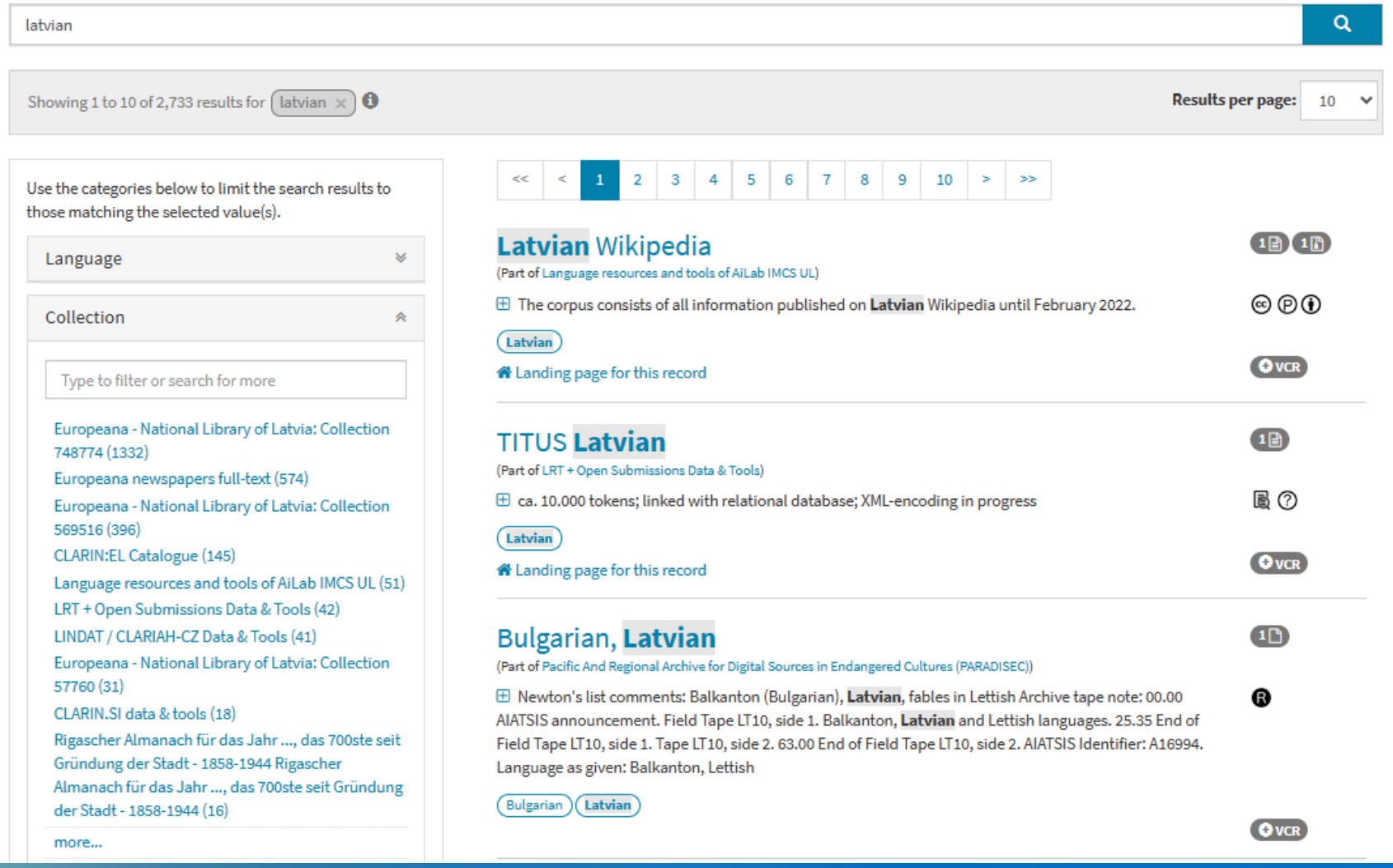
CLARIN-LV

Latviešu valodas resursi citu valstu CLARIN repozitorijos





CLARIN virtuālā valodas resursu krātuve (VLO)



https://vlo.clarin.eu

CLARIN valodas resursu kopas

Corpora

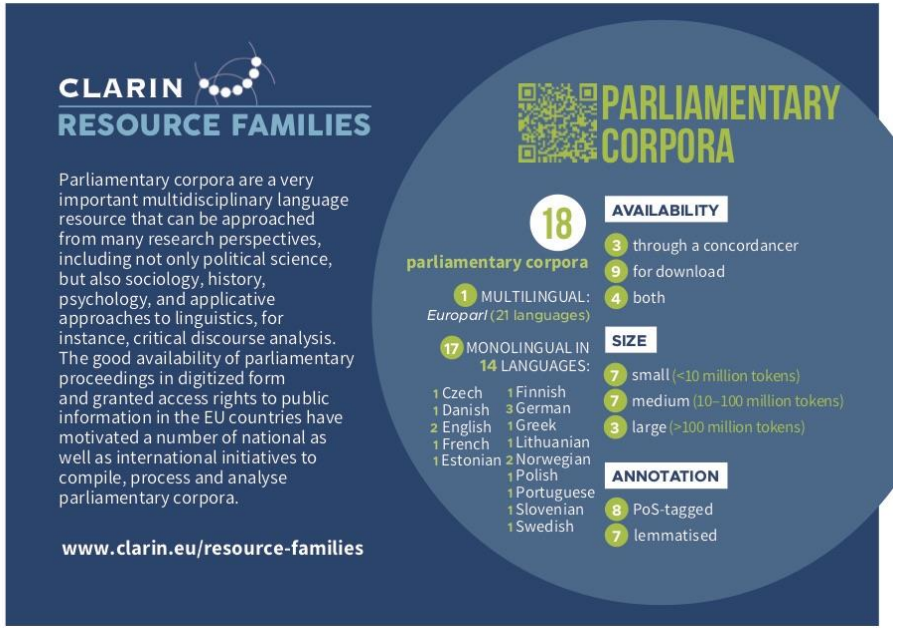
- Computer-Mediated Communication Corpora
- Corpora of Academic Texts
- Historical Corpora
- L2 Learner Corpora
- Legal Corpora
- Literary Corpora
- Manually Annotated Corpora
- Multimodal Corpora
- Newspaper Corpora
- Oral History Corpora
- Parallel Corpora
- Parliamentary Corpora
- Reference Corpora
- Sign Language Resources
- Spoken Corpora

Lexical Resources

- Language Models
- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

Tools

- Corpus Query Tools
- Normalisation
- Named Entity Recognition
- Part-of-Speech Tagging and Lemmatisation
- Tools for Sentiment Analysis



Resource Families | CLARIN ERIC

CLARIN valodas resursu kopas

Corpora

- Computer-Mediated Communication Corpora
- Corpora of Academic Texts
- Historical Corpora
- L2 Learner Corpora
- Legal Corpora
- Literary Corpora
- Manually Annotated Corpora
- Multimodal Corpora
- Newspaper Corpora
- Oral History Corpora
- Parallel Corpora
- Parliamentary Corpora
- Reference Corpora
- Sign Language Resources
- Spoken Corpora

Lexical Resources

- Language Models
- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries

Tools

- Corpus Query Tools
- Normalisation
- Named Entity Recognition
- Part-of-Speech Tagging and Lemmatisation
- Tools for Sentiment Analysis

Corpora of Academic Texts

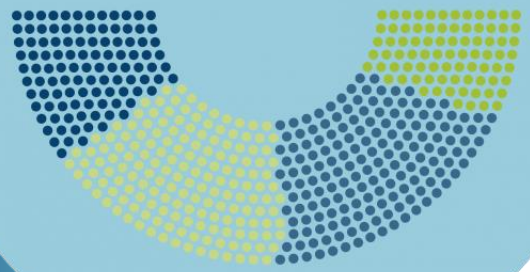
Corpora of academic texts contain scholarly writing, such as research papers, essays and abstracts published in academic journals, conference proceedings, and edited volumes, theses written by students at undergraduate and graduate levels, and scientific monographs.

The CLARIN ERIC infrastructure gives access to 24 corpora of academic texts, 2 of which are multilingual and 22 monolingual. The available corpora contain scholarly texts in the following 11 languages: Czech, English, Estonian, Finnish, French, German, Greek, Russian, Slovenian, Spanish, and Swedish. More than 15 different scholarly disciplines are represented, with the most prominent being linguistics, computer science, economics, and medicine. The majority of the corpora are richly tagged and are available under public licences.

Resource Families | CLARIN ERIC

Sadarbība, veidojot kopīgus valodas resursus

ParlaMint



Corpus

CLARIN.SI Data & Tools

Multilingual comparable corpora of parliamentary debates ParlaMint 2.0

(CLARIN ERIC / 2021-05-10)

Author(s):

Erjavec, Tomaž ; Ogrodniczuk, Maciej ; Osenova, Petya ; Ljubešić, Nikola ; Simov, Kiril ; Grigorova, Vladislava ; Rudolf, Michał ; Pančur, Andrej ; Kopp, Matyáš ; Barkarson, Starkaður ; Steingrímsson, Steinþór ; van der Pol, Henk ; Depoorter, Griet ; de Does, Jesse ; Jongejan, Bart ; Haltrup Hansen, Dorte ; Navarretta, Costanza ; Calzada Pérez, María ; de Macedo, Luciana D. ; van Heusden, Ruben ; Marx, Maarten ; Çöltekin, Çağrı ; Coole, Matthew ; Agnoloni, Tommaso ; Frontini, Francesca ; Montemagni, Simonetta ; Quochi, Valeria ; Venturi, Giulia ; Ruisi, Manuela ; Marchetti, Carlo ; Battistoni, Roberto ; Sebők, Miklós ; Ring, Orsolya ; Darģis, Roberts ; Utka, Andrius ; Petkevičius, Mindaugas ; Briedienė, Monika ; Krilavičius, Tomas ; Morkevičius, Vaidas

This item contains 17 files (2.03 GB).

Publicly Available

CLARIN

ParlaMint

Showcase 2: 'COVID' keywords

Italy	Poland	Slovenia	UK
pandemic	pandemic	epidemic	covid
covid	*bell*	ventilator	coronavirus
covid-19	coronavirus	coronavirus	**furlough**
coronavirus	covid-19	covid-19	lockdown
lockdown	mask	virus	pandemic
mask	epidemic	quarantine	distancing
recovery	quarantine	corona	**ppe**
European Stability Mechanism (ESM)	Kukiz15	mask	*inaudible*
distancing	the Left	pandemic	
fund	**shield**	**anti-corona (adj)**	
serologic	**anti-crisis**	/paramilitary group/	
virus	covid	/name/	

Parliamentary Debates in the COVID Times

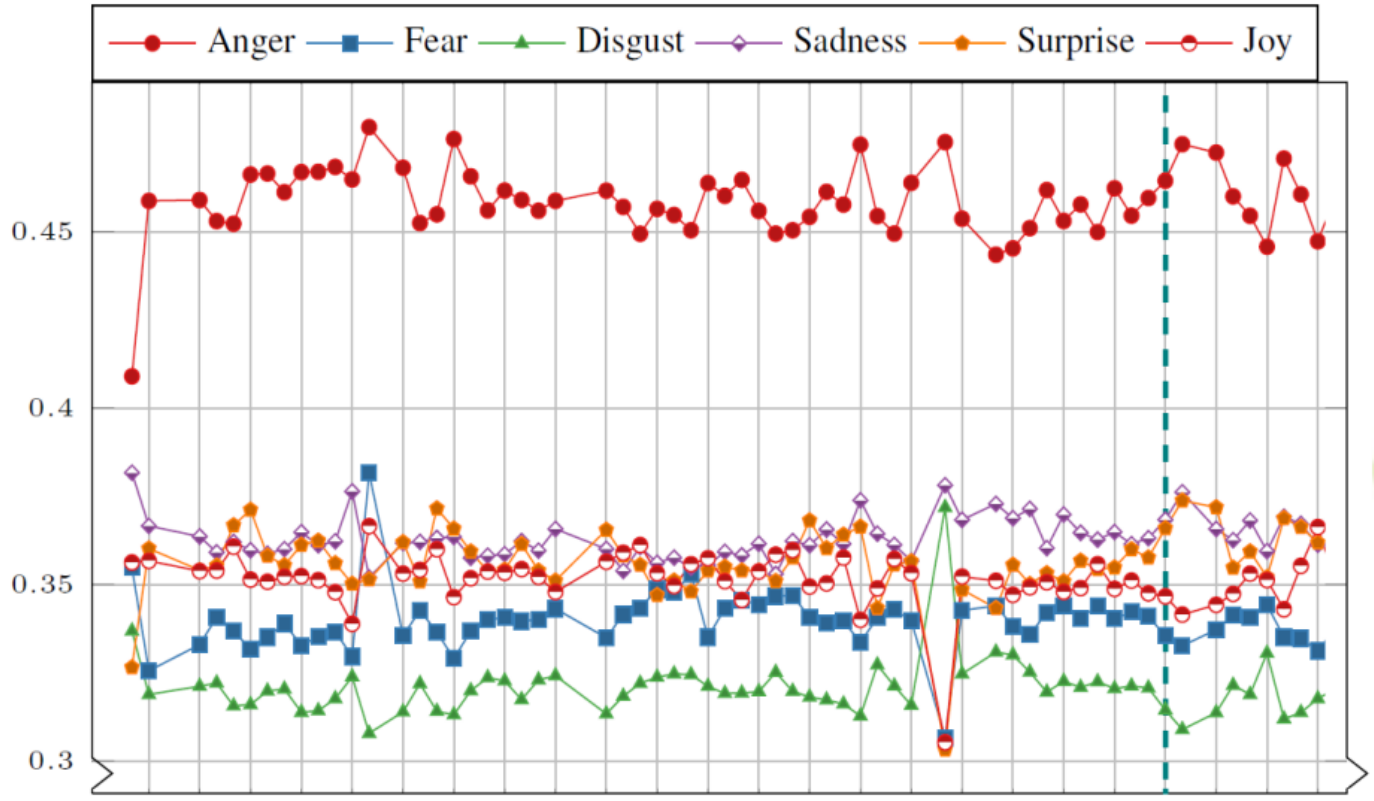
Helsinki DH Hackathon 21

CLARIN

Showcase 6: Emotions Running High

Kurtoğlu and Çöltekin (2022). Emotions Running High?

- Investigating polarization of politics by assigning emotion scores to speeches
- anger being the dominant emotion
- the ruling party showing more stable emotions compared to the opposition



CLARIN zināšanu centrs morfoloģiski bagātām valodām SAFMORIL

- SAFMORIL ir CLARIN zināšanu centrs (knowledge centre) morfoloģiski bagātām valodām
- SAFMORIL konsultē digitālo humanitāro zinātņu pētniekus un valodu tehnoloģiju izstrādātāju par valodas resursiem un rīkiem morfoloģiski bagātām valodām
- SAFMORIL apvienojušās Helsinku Universitāte, Trumses Universitāte, Latvijas Universitātes Matemātikas un Informātikas institūts un Vītauta Dižā Universitāte

CLARIN K CENTRE

CLARIN-LV | Par CLARIN | Konsorcijs | Repozitorijs | Konferences un semināri | Jaunumi | Kontakti

Konference "Jaunākie valodas resursi un rīki digitālajām humanitārajām zinātnēm"

2023. gada 12. septembrī Latvijas Universitātes Matemātikas un informātikas institūts sadarbībā ar digitalhumanities.lv un DHELI projektu rīkoja ikgadējo CLARIN-LV konferenci "Jaunākie valodas resursi un rīki digitālajām humanitārajām zinātnēm". Konferences materiāli:

- *Kaspars Bērziņš (VPC).* Augstākās izglītības un zinātnes IT koplietošanas pakalpojumu centrs (VPC) un datu kuratoru projekts
- *Maciej Ogrodniczuk* (CLARIN ERIC). ParlaMint: Multilingual Comparable Corpora of Parliamentary Debates for Digital Humanities
- *Edward Gray* (DARIAH EU). The Digital Research Infrastructure for the Arts and Humanities: How DARIAH Helps Enable SSH Researchers
- *Inguna Skadiņa* (LU MII). CLARIN-LV - solis pretī valodu digitālai vienlīdzībai
- *Anta Trumpa* un *Sanda Rapa* (LU LaVI). Valodas digitalizācija: no pirmajiem rakstu avotiem līdz mūsdienām
- *Anda Baklāne* (LNB). Digitālo pakalpojumu aktualitātes Latvijas Nacionālajā bibliotēkā
- *Antra Kļavinska* (RTA). Latgaliešu valodas digitālie resursi: paveiktais, klupšanas akmeņi un iespējas
- *Sanita Reinsone* un *Sandis Laime* (LU LFMI). Humma.lv – ceļā uz digitālo platformu humanitārajām un mākslas zinātnēm
- *Valts Ernštreits* (LU Lībiešu institūts). Lībiešu valodas datu ieguve un tās nākotnes perspektīvas
- *Baiba Saulīte un Ilze Auziņa* (LU MII). Tēzaurs.lv un Korpuss.lv jaunumi

Seminārs "Runas korpusu izveide un to izmantojums"

2023. gada 18. maijā notika *digitalhumanities.lv* seminārs sadarbībā ar Clarin-LV un DARIAH-EU "Runas korpusu izveide un to izmantojums". Semināra materiāli:

- Ilze Auziņa. Runas korpusu nozīme valodu tehnoloģiju izstrādē.
- Antra Kļavinska. Mūsdienu latgaliešu runas korpusa izveide mazāk lietoto valodu dokumentēšanas kontekstā.
- Ilze Auziņa un Guna Rābante-Buša. Runas datu apstrāde programmā ELAN.

CLARIN K CENTRE

Citi būtiski CLARIN piedāvājumi

Learning Hub

- Learning and Training Resources
- Training and Workshops
- Digital Humanities Course Registry
- Notebooks in Education

Learning Hub | CLARIN ERIC

Funding Hub

- Workshop Funding
- User Involvement Funding
- Teaching with CLARIN
- Trainer Network Programme
- Mobility Grants
- Resource Families Project Funding

CLARIN Funding Hub | CLARIN ERIC

CLARIN konference

- Iespēja piedalīties ikvienam konsorcija partnerim (atbalsts referātiem)
- Īpašs atbalsts doktorantiem

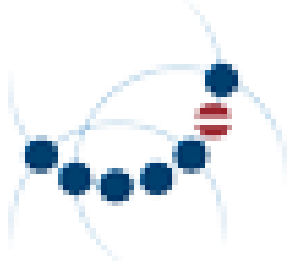
Konsorcijs

LU LITERATŪRAS, FOLKLORAS UN MĀKSLAS INSTITŪTS

Latvijas Universitātes
Matemātikas un informātikas institūts

L N B
LATVIJAS NACIONĀLĀ BIBLIOTĒKA

CLARIN-LV



LIEPĀJAS UNIVERSITĀTE

LATVIJAS UNIVERSITĀTE

RĪGAS STRADIŅA UNIVERSITĀTE

Paldies par uzmanību!

VPP-LETONIKA-2021/1-0006

2.3.1.1.i.0/1/22/I/CFLA/002