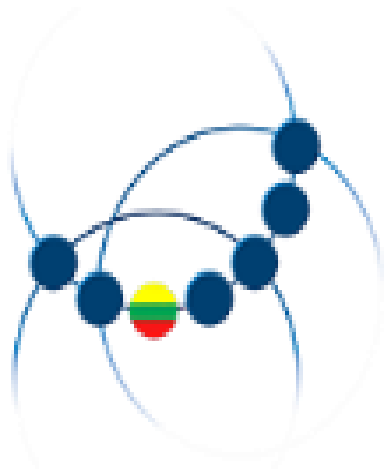


CLARIN-LT



CLARIN-LT konsorciumo veikla ir planai

Andrius Utkas

Kompiuterinės lingvistikos centras /
CLARIN centras

Pranešimo planas

1. Svarbiausi 2016 m. įvykiai CLARIN ERIC infrastruktūroje;
2. Svarbiausi įvykiai CLARIN-LT konsorciume;
3. Atlikti darbai;
4. Ateities planai.

CLARIN ERIC: Kas pasikeitė per metus?



- Per 2016 metus prie 17 (Austrija, Bulgarija, Čekija, Danija, Estija, Graikija, Italija, Lenkija, Lietuva, Nyderlandai, Norvegija, Olandų kalbos sąjunga, Portugalija, Slovėnija, Suomija, Švedija, Vokietija) CLARIN ERIC narių prisijungė dar 3 šalys.
- Naujieji nariai tai: Vengrija ir Latvija, bei stebėtojo statusą įgijusi Didžioji Britanija.
- Sertifikuotų centrų skaičius išaugo nuo 17 iki 20.
- Dar keletas centrų ruošiasi sertifikavimui (tarp jų ir CLARIN-LT centras).

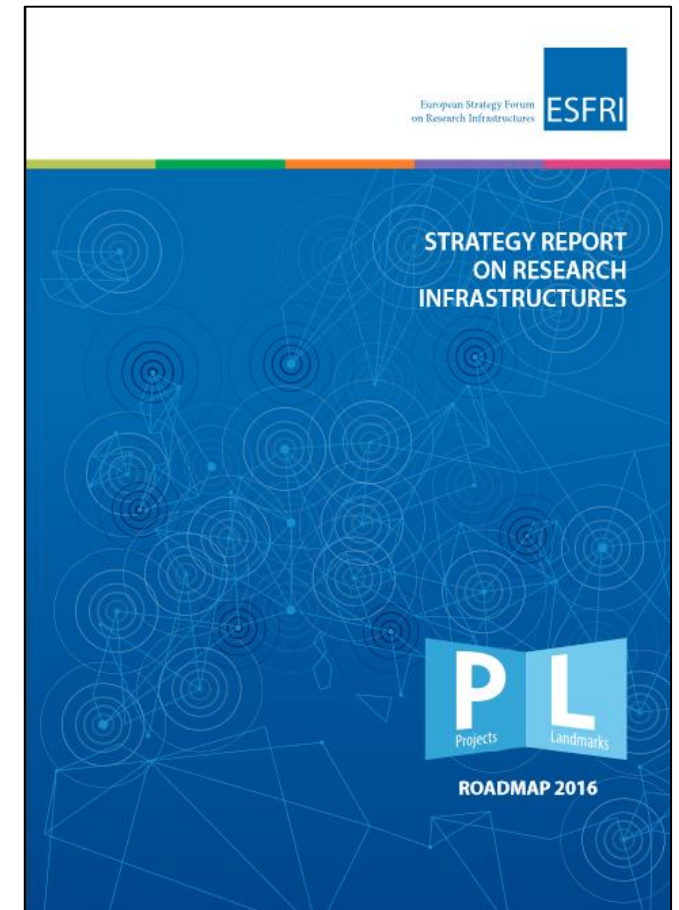
CLARIN ERIC centrai



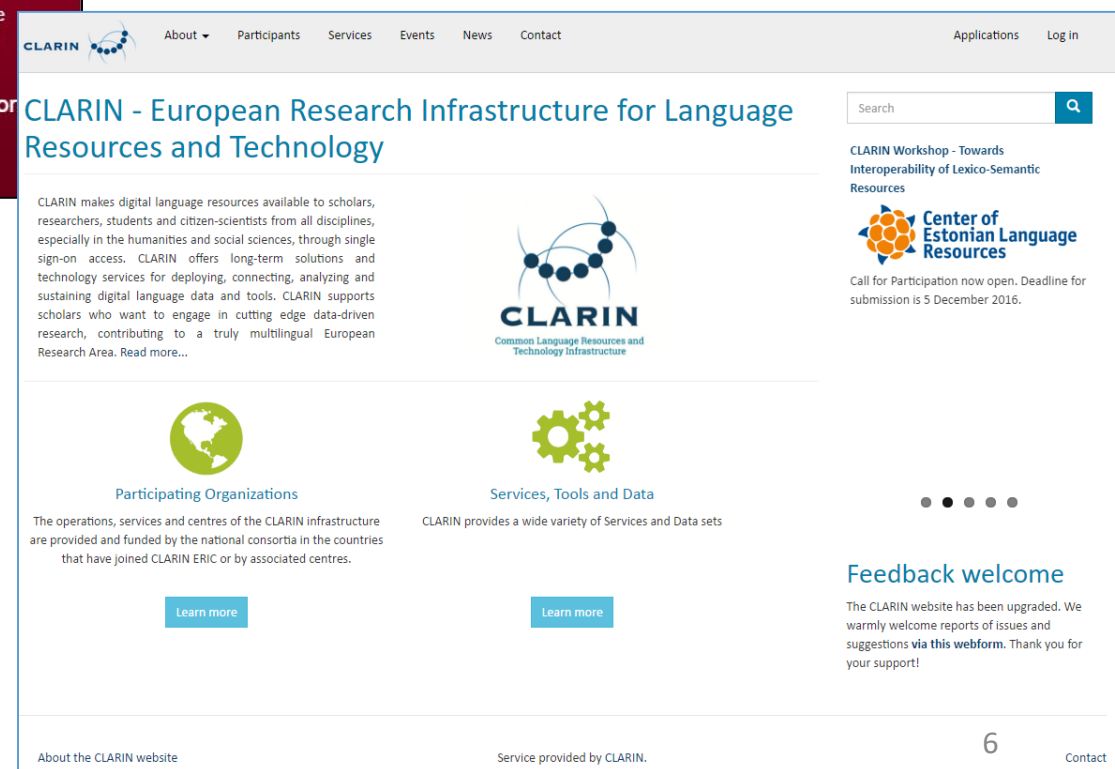
- 1
- 2
- 3
- 4

CLARIN ERIC tarp svarbiausių ESFRI infrastruktūrų

- 2016 m. kovo mėn. Europos mokslinių infrastruktūrų strategijos forumas (ESFRI) paskelbė „ESFRI gaires“ (Roadmap 2016);
- Gairėse buvo oficialiai patvirtintos 29 infrastruktūros, kurios įvardijamos kaip atitinkančios ilgalaikius Europos mokslinių tyrimų interesus (angl. *landmark*);
- CLARIN ERIC yra viena iš tokių infrastruktūrų;
- Humanitarinių ir socialinių mokslų srityje dar įvardintos 4 infrastruktūros: DARIAH, CESSDA, ESS, SHARE.



Pasikeitė CLARIN puslapio dizainas



CLARIN ERIC dalyviai LREC 2016 konferencijoje

- 2016 m. gegužės mėnesį CLARIN gausiai dalyvavo LREC'2016 konferencijoje Portorože (Slovėnija);
- CLARIN-LT atstovavo Tomas Krilavičius ir Andrius Utkas;
- Pristatėme 2 straipsnius ir kartu su kolegomis iš Ciuricho organizavome seminarą „Normalisation of Social Media Texts“;
- CLARIN nariai pristatė 40 straipsnių ir organizavo 7 konferencijos seminarus.



Pasikeitė konsorciumo lyderis



Steven Krauwer



Franciska de Jong

Pasikeitė direktorių taryba



Franciska de Jong

Executive Director CLARIN ERIC



Bente Maegaard

Vice Executive Director



Koenraad De Smedt

Chair of the NCF



Dieter Van Uytvanck

Chair Centre Committee & Technical Director



Darja Fišer

Director of User Involvement

Pasikeitė CLARIN Generalinės Asamblėjos pirmininkas



Helge Kahler (Vokietija)



Rūta Petrauskaitė (Lietuva)

Pasikeitė CLARIN-LT nacionalinis koordinatorius



Rūta Petrauskaitė



Jurgita Vaičenonienė

Svarbiausi CLARIN-LT įvykiai

- CLARIN-LT atstovai dalyvavo LREC'2016 konferencijoje;
- Gintarė Grigonytė buvo nominuota „Steven Krauwer Award for Young Scientists“.
- Net 9 žmonės iš CLARIN-LT dalyvavo metinėje CLARIN konferencijoje Aix en Provance (Prancūzijoje);
- Dalyvavome rengiant COST paraišką apie socialinių tekstų tekstynus.



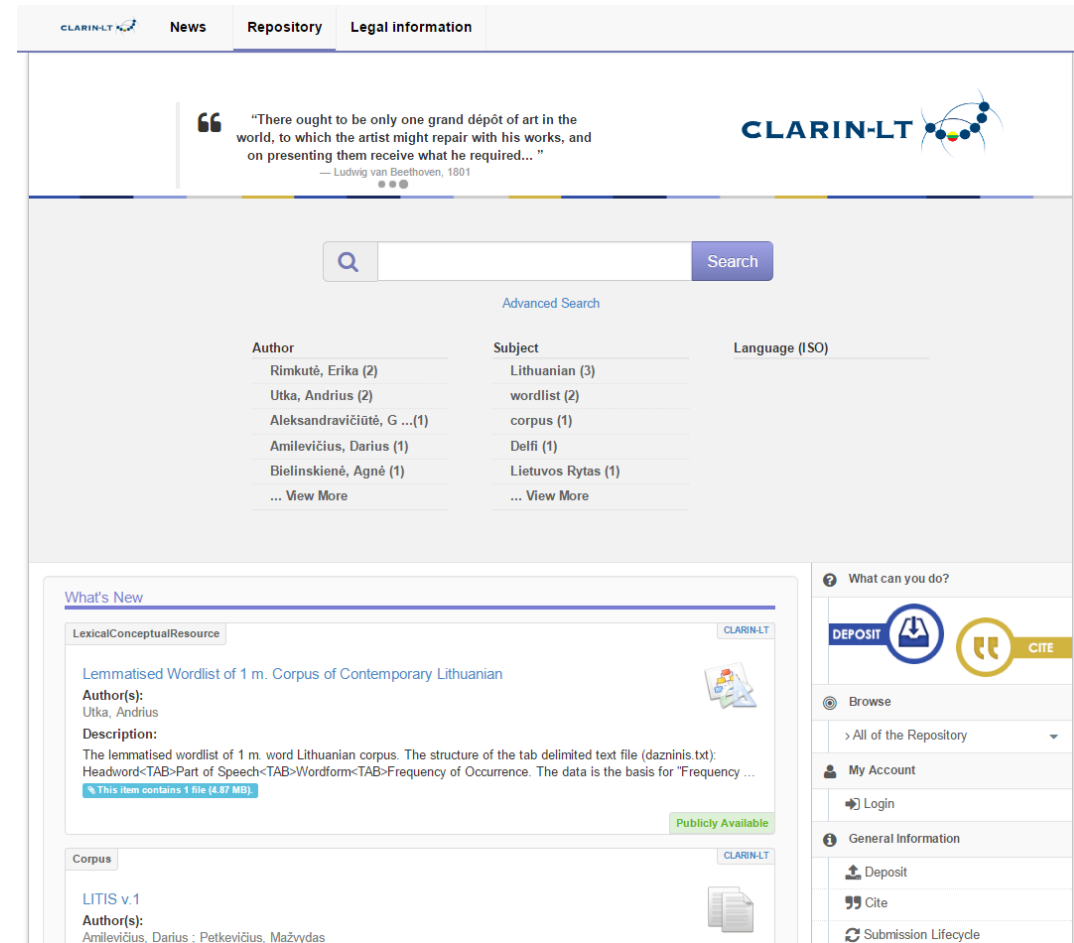
Atlikti darbai: saugykla (1)

Jau veikia ir pirmuosius 5 išteklius saugo CLARIN-LT saugykla:

1. ALKSNIS – sintaksinių medžių tekstynas (PUB);
2. MATAS – morfologiškai anototas tekstynas (ACA);
3. LITIS – socialinis tekstų tekstynas (ACA);
4. 1 mln. žodžių lemuotas dažninis sąrašas (PUB);
5. DLKT žodžių formų dažninis sąrašas (PUB).

Ne vis darbai su saugykla dar atlikti:

- Pradėtas procesas CLARIN-LT įstojimo į CLARIN paslaugų tiekėjų federaciją (SPF),
- Duomenų pripažinimo antspaudas – (*Data Seal of Approval*).



Atlikti darbai: licencijos

- Saugykloje jau įdiegtas mechanizmas kaip priskirti CLARIN-LT licencijas, kurios šiuo metu gali būti 3 rūšių:
 - PUB
 - ACA
 - RES
- Licencijos yra suderintos su Lietuvos Respublikos įstatymais ir pateikiamos anglų ir lietuvių kalba.
- Visa susijusi teisinė informacija jau pateikta saugykloje.

Read and accept the [Distribution License Agreement](#)

Accepted ☒

By checking this box, you agree to the Distribution License Agreement for this repository to reproduce, translate and distribute your submissions worldwide.

 If you have questions regarding this licence please contact the [Help Desk](#).

Select the resource license



The License Selector will provide you visual assistance to select the most appropriate license for your data or software. For the list of all supported licenses and their details visit [License List Page](#).

OPEN License Selector

- OR -

If you already know under which license you want to distribute your work, please select from the dropdown below.

PUB CLARIN-LT-Distribution_Licence_EN-LT

PUB PUB_CLARIN-LT_End-User-Licence-Agreement_EN-LT

PUB CLARIN-LT_Privacy_Policy_EN-LT

PUB CLARIN-LT_Terms-of-Service_EN-LT

PUB CLARIN-LT-Distribution_Licence_EN-LT

ACA ACA_CLARIN-LT_End-User-Licence-Agreement_EN-LT

RES RES_CLARIN-LT_End-User-Licence-Agreement_EN-LT

< Previous

Save & Exit

Next >

Atlikti darbai: ištekliai

- Naudojant automatinius paieškos serverius semantika.lt serveriuose toliau kaupiamas bendrasis internetinis tekstynas (BIT) (2015-2016 m. ~ 400 mln. žodžių);
- Sukauptas sintaksiškai anototas tekstynas (2 355 sakiniai, ~30 tūkst. žodžių) – ALKSNIS;
- Toliau kaupiamas lietuvių kalbos tarmių garsynas (pietų ir rytų aukštaičių tarmės). Sukaupta 15 val.;
- 20 tūkst. sinsetų papildyta lietuvių kalbos leksinė semantinė ontologija *LitWordNet*;
- Sukurtas socialinių tekstų komentarų tekstynas LITIS. Šiuo metu tekстыne 199 909 komentarai.
- Toliau buvo tobulinama lietuvių kalbos morfologinė duomenų bazė.

Atlikti darbai: paslaugos

Paieška tekstyne Išplėstinė paieška

Forma	Lema	Morfologija	?	leškoti	
-------	------	-------------	---	---------	--

TEKSTYNAS

Visas tekstynas

RUSIAVIMAS

Nerūšiuoti	-3	-2	-1	0	1	2	3
------------	----	----	----	---	---	---	---

- Sukurta ir toliau tobulinama <http://corpus.vdu.lt> paieška bei DLKT tekstyno saugykla;
- Patobulinta lietuvių k. valdymo balsu internetinė paslauga (asmenų vardų ir vietovardžių atpažinimas);
- Tobulinama prieiga prie semantinės ontologijos *LitWordNet*. Prieiga: <http://mackus.vdu.lt/LitWordNet/>

Atlikti darbai: įkurti 2 informacijos centrai

Lietuviškai English

- Vykdydami savo įsipareigojimus kurti „Žinių dalinimosi infrastruktūrą“, VDU ir KLC įsteigė 2 informacijos centrus
 - Lietuvių kalbos technologijų
 - Semantinio koncepcinio modeliavimo
- Informacija apie šiuos centrus pateikiama clarin-lt.lt svetainėje;
- Abu partneriai įsipareigoja nemokamai konsultuoti visuomenę lietuvių kalbos technologijų ir semantinio koncepcinio modeliavimo klausimais.



CLARIN-LT

Naujienos Paslaugos Apie CLARIN ERIC Saugykla Partneriai Kontaktai DUK Informacijos centras Renginiai

In Archive

- 2016 m. lapkričio mėn.
- 2016 m. spalio mėn.
- 2016 m. rugsėjo mėn.
- 2016 m. liepos mėn.
- 2016 m. birželio mėn.
- 2016 m. gegužės mėn.
- 2016 m. balandžio mėn.
- 2016 m. vasario mėn.
- 2016 m. sausio mėn.
- 2015 m. gruodžio mėn.
- 2015 m. spalio mėn.

Informacijos centras

Lietuvių kalbos technologijų informacijos centras Vytauto Didžiojo universitetas (Kaunas)	Aprašymas: VDU KLC ir VDU IF turi didelę patirtį lietuvių kalbos tekstų lingvistikos ir natūralios kalbos apdorojimo srityse. Informacijos centras teikia informaciją ir konsultuoja šiomis temomis: • morfologija, sintaksė ir semantika; • morfologinis anotavimas ir automatinė sintaksinė analizė; • tekstų analizė, terminų ištraukimas, automatizuotas vertimas, socialinių tekstų apdorojimas, šnekos atpažinimas ir kt. Ilgiausia atsakymo trukmė: 3 darbo dienos. El. paštas: info.eta.clarin.vdu.lt Telefonas: +37037327841
Semantinio ir koncepcinio modeliavimo informacijos centras Kauno Technologijos universitetas (Kaunas)	Aprašymas: KTU Informatikos fakulteto Informacijos sistemų katedra turi nemažai kompetencijos duomenų bazių, žinių ir informacinių sistemų inžinerijos, koncepcinio ir semantinio modeliavimo, ontologijų kūrimo, semantinės paieškos taikymo srityse. Konsultacijų centras teikia informaciją ir konsultuoja šiais klausimais: • veiklos žodynų specifikuojamas ribota natūralia kalba, taikant SBVR (angl. Semantic of Business Vocabulary and Business Rules) standartą, • veiklos žodynų ir taisyklių taikymas; • semantinės paieškos (prasmės informacijos ištraukimas); • teksto apdorojimas semantine paieška (informacijos ištraukimas), įskaitant natūralios kalbos apdorojimo metodus, semantinį anotavimą ir ontologijų pildymą; • semantinės paieškos sąsajos natūralia kalba. Ilgiausia atsakymo trukmė: 3 darbo dienos. El. paštas: rimantas.butleris.eta.ktu.lt Telefonas: +37037453445

Ateities planai: finansavimas

- Sekančiam periodui infrastruktūroms planuojama skirti ES struktūrinę paramą. Jau žinoma, kad bus finansuojami:
 - CLARIN-LT narystės mokestis;
 - Įrenginiai ir technika.
- Tiesa, veiklų kol kas neplanuojama finansuoti.
- Dėl Lietuvos valdžios pasikeitimo, infrastruktūrų finansavimo klausimai vis atidedami todėl dar nepavyko užsitikrinti finansavimo sekančiam periodui.

Ateities planai: darbai

CLARIN ERIC

2017 metais ir toliau bus ypatingai akcentuojamas naujų vartotojų pritraukimas.

CLARIN LT

Jei pavyktų gauti finansavimą kalbos išteklių vystymui, tai planuotume toliau vystyti šiuos išteklius:

- ALKSNĮ;
- Lietuvių kalbos tarmių garsyną;
- Socialinių tekstų tekstyną;
- Saugykloje saugojimui paruoštume egzistuojančius išteklius (projekto ASTRA tekstynus, anaforiškai anototą tekstyną ir kt.)

Ateities planai: renginiai

CLARIN ERIC

- 2017 m. rugsėjo 18-20 d. metinė konferencija vyks Budapešte;

CLARIN-LT

- 2017 m. gegužę, CLARIN-LT organizuoja tarptautinį seminarą Kaune

“Creation and Use of Social Media Resources”



Ačiū už dėmesį!