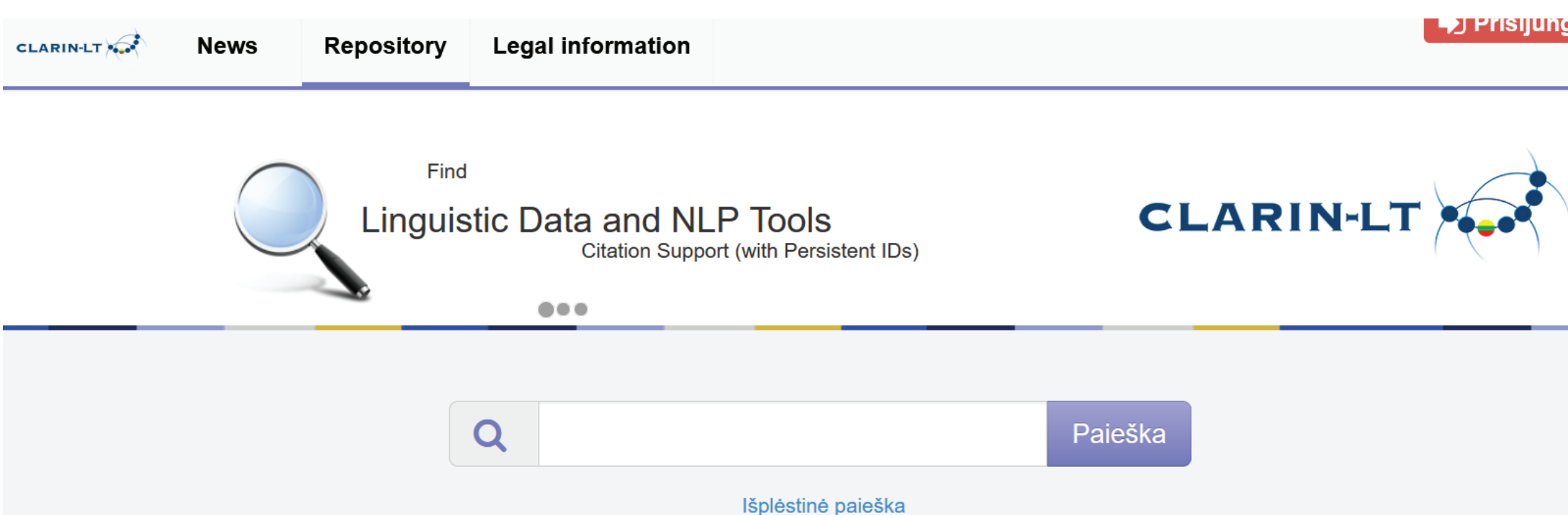


Lietuvių kalbos tekstynai atvirosios prieigos saugykloje CLARIN-LT

ERIKA RIMKUTĖ | JOLANTA KOVALEVSKAITĖ | JURGITA VAIČENONIENĖ

Vytauto Didžiojo universiteto Skaitmeninių išteklių ir tarpdisciplininių tyrimų institutas

Nacionalinė CLARIN-LT saugykla skirta humanitarinių ir socialinių mokslų tyrėjų duomenims saugoti. Duomenų saugykla – tai tarsi kalbos duomenų ir teksto analizės įrankių biblioteka. Saugykloje galima ieškoti duomenų ir įrankių, lengvai juos atsisiųsti arba įkelti savo duomenis, kad jie būtų saugiai saugomi, juos galėtų rasti, naudoti ir teisingai cituoti kiti tyrėjai. Duomenims suteikiamas unikalus ir nuolatinis skaitmeninis objekto identifikatorius, jie tampa prieinami Lietuvos ir užsienio tyrėjams.



Vytauto Didžiojo universiteto Skaitmeninių išteklių ir tarpdisciplininių tyrimų institute parengti ir (ar) atvirosios prieigos saugykloje CLARIN-LT prieinami tekstynai

Šiuo metu saugykloje yra prieinami 57 lietuvių kalbos ištekliai ir įrankiai. Išteklių grupėje daugiausia yra tekstynų, tačiau galima rasti ir išvestinių tekstynų produktų, pvz., dažninių sąrašų, n-gramų sąrašų, taip pat įrankių, duomenų bazių. Ištekliai gali būti įkeliami priskiriant jiems atvirosios, akademinės arba ribotos prieigos licenciją.

- 1) Bendro pobūdžio:** „Dabartinės lietuvių kalbos tekstynas“ (<http://tekstynas.vdu.lt/tekstynas/> arba <http://corpus.vdu.lt/lt/>)
- 2) Lygiagretieji:**
 - anglų–lietuvių–anglų, čekų–lietuvių–čekų (<https://sitti.vdu.lt/lygiagretus-tekstynas/>)
 - latvių–lietuvių–latvių (<https://sitti.vdu.lt/lila-lygiagretusis-tekstynas/>)
- 3) Specializuoti:**
 - anglų–lietuvių lygiagretusis tekstynas DVITAS (kibernetinio saugumo terminų tekstynas <https://hdl.handle.net/20.500.12259/240110>)
 - anglų–lietuvių palyginamasis tekstynas DVITAS (<https://hdl.handle.net/20.500.12259/240304>)
 - Nusikaltimo diskurso tekstynas (<https://hdl.handle.net/20.500.12259/240288>)
 - DIGIRES COVID-19 tekstynas (patikimi ir nepatikimi žiniasklaidos straipsniai apie COVID-19 pandemiją, <https://hdl.handle.net/20.500.12259/252147>)
 - Lietuvos Respublikos Seimo stenogramų tekstynas (<https://hdl.handle.net/20.500.12259/240256>)
 - KLASIUS (900 ištraukų iš vidurinių mokyklų moksleiviams skirtų vadovėlių ir leidinių, kurie įtraukti į privalomų universiteto studijų programų bibliografinių šaltinių sąrašus, <https://hdl.handle.net/20.500.12259/242418>)
 - LITIS tekstynas (portalų www.delfi.lt, www.lrytas.lt komentarų tekstynas <https://hdl.handle.net/20.500.12259/240308>)
 - LITUND tekstynas (tikrovės neatitinkantys naujienų tekstai, LRT.lt portalo panašios tematikos tekstai, <https://hdl.handle.net/20.500.12259/280548>)
 - „Mokomasis lietuvių kalbos tekstynas“ (<https://kalbu.vdu.lt/mokymosi-priemones/mokomasis-tekstynas/>)
 - „Lietuvių kalbos mokinių tekstynas“ (<https://kalbu.vdu.lt/mokymosi-priemones/mokiniu-tekstynas/>)
- 4) Vertimo:** ORVELIT tekstynas (<https://hdl.handle.net/20.500.12259/240292>)
- 5) Anotuoti** (kai kurių yra kelios versijos):
 - morfologiškai anotuotas tekstynas MATAS (<http://hdl.handle.net/20.500.11821/61>)
 - sintaksiškai anotuotas tekstynas ALKSNIS (<http://hdl.handle.net/20.500.11821/21>)
 - šiuo metu rengiami morfologiškai ir sintaksiškai anotuoti tekstynai: (<https://sitti.vdu.lt/morfologiskai-ir-sintaksiskai-anotuotu-tekstynu-modeliai-dirbtinio-intelektu-apmokymui/>)
- 6) Sakytinės kalbos tekstynai, garsynai:**
 - „Sakytinės lietuvių kalbos tekstynas“ (<http://sakytinistekstynas.vdu.lt/index.php>)
 - Skaitomo teksto garsynas 7G (<https://hdl.handle.net/20.500.12259/269385>)

Projektai

2024–2029 m. CLARIN-LT konsorciumas įgyvendina „Narystės tarptautinėje mokslinių tyrimų infrastruktūroje CLARIN ERIC (Bendrųjų kalbos išteklių ir technologijų infrastruktūra) planą“.



2024–2026 m. SITTI įgyvendina Europos Sąjungos NextGenerationEU projektą „Morfologiškai ir sintaksiškai anotuotų tekstynų modeliai apmokymui (auksiniai standartai)“ (Nr. 02-098-K-0001).



NAUJOS KARTOS
LIETUVA



Finansuoja
Europos Sąjunga
NextGenerationEU