

Riikliku programmi „Eesti keeletehnoloogia 2011–2017” (EKT, vt www.keeletehnoloogia.ee) eesmärk on saavutada Eestis keeletehnoloogiline tase, mis võimaldab eesti keelel edukalt toimida tänapäeva infotehnoloogilises maailmas. Programm rahastab keeletehnoloogilist teadus- ja arendustegevust alates ressursside loomisest kuni keeletehnoloogiliste rakenduste prototüüpide loomiseni. Uuri, kas ka Sinu idee võiks programmist rahastust leida!



CLARIN (*Common Language Resources and Technology Infrastructure*) ERIC (*European Research Infrastructure Consortium*) on Euroopa riikide teaduskonsortsium, mille eesmärk on teha Euroopa erinevate keelte digiressursid lihtsalt kättesaadavaks ja kasutatavaks. Uuri lähemalt www.clarin.eu.



META-NET (vt www.meta-net.eu) on projektide võrgustik, mille eesmärk on edendada mitmekeelse Euroopa infoühiskonna tehnoloogilisi aluseid. META-SHARE on avatud, turvaline, koostalitlusvõimeline hajusinfrastruktuur keeleressursside jagamiseks ja vahetamiseks, kuhu panustavad oma keeletehnoloogiliste ressurssidega kõik META-NETi partnerid. EKRR ühendub META-SHARE võrgustikku oma sõlme (vt metashare.ut.ee) kaudu.



Eesti Keeleressursside Keskus
J. Liivi 2–340, 50409, Tartu, Eesti
Telefon: 7 376 433
E-post: info@keeleressursid.ee



Kadri Vider
EKRR tegevjuht
CLARIN ERIC riiklik koordinaator Eestis
Telefon: 7 376 433
E-post: kadri.vider@ut.ee



Tiit Roosmaa
EKRR nõukogu esimees
E-post: tiit.roosmaa@ut.ee

Jaga ja kasuta!

keeleressursid.ee



*Magusad murakad
eesti keele
digivara korvis*

KEELERESSURSSIDE KESKUS

Eesti Keeleressursside Keskus (EKRK) on teadustaristu, mis teeb huvilistele kättesaadavaks eesti keele digiressursid ja -tehnoloogia.

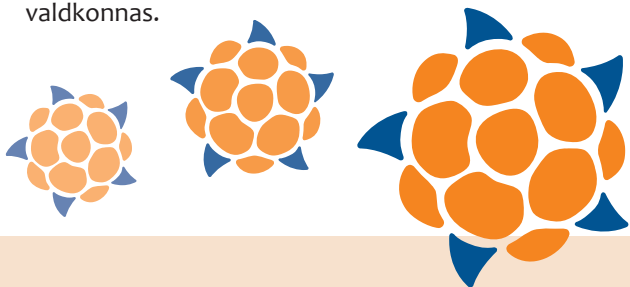
Keskuse tegevuse eesmärk on koondada olemasolevad digitaalsed keeleressursid (sõnastikud, teksti- ja kõnekorpused, keeleandmebaasid) ja keele töötlemise vahendid (tarkvara) vastastikku toimivaiks ning oskusteabega varustatud teenusteks, mida kasutajad saavad vajaduse korral ka oma tarbeks kohandada. Kõik keeleressursid peavad vastama rahvusvahelistele standarditele ja olema varustatud kasutuslitsentsiga. Eesti-siseselt on taristu avatud nii keeleressursside omanikele, arendajatele kui ka kasutajatele, kes nõustuvad kasutus- ja litsentsitingimustega. Eelistatud on teaduskasutajad ning eesti keele kasutajaskonnale avalikes huvides pakutavad keeleressursid ja keeletehnoloogilised vahendid.

MIDA EKRK PAKUB?

Digiressurssi saab kasutada korduvalt, kuid selleks on vaja neid koguda, arhiveerida ja kasutajatele kättesaadavaks teha.

Kasutuse hõlbustamiseks koondatakse juurdepääs olemasolevatele digitaalsetele arhiividele ja pakutakse keeletehnoloogia vahendeid kui veebiteenust, mis kasutab arhiveeritud andmeid. EKRK juhendab ressursside dokumenteerimist ja arhiveerimist ning arendab vajalikke keeletehnoloogia standardeid, koostab lepinguid ja varustab digiressurssi erinevat tüüpi (nt üldine avalik, akadeemiline ja ärikasutus) kasutuslitsentsidega.

EKRK tegeleb ka keeleressursside ja nende kasutusvõimaluste tutvustamisega oma võimalikele sihtgruppidele eeskätt sotsiaal- ja humanitaarteaduste valdkonnas.



KEELERESSURSID

Tekstikorpused

EKRK haldab nii üldisi (nt eesti keele baaskorpus) kui ka valdkonnapõhiseid korpuseid (nt jututubade tekstid); nii ükskeelseid kui ka paralleelkorpusi, mis sisaldavad tõlkeid mitmes keeles. Algtekstile lisaks võib korpus sisaldada eri tasandite märgendust (nt morfoloogia, süntaks, semantika, isikunimed).

- 🌸 Eesti keele koondkorpus
- 🌸 Segakorpus: Uus meedia
- 🌸 Eesti keele puudepank

Kõneandmebaasid

Kõnekorpused, mis sisaldavad helisalvestisi, vahel koos transkribeeritud tekstiga; kõne sünteesiks vajalikud andmebaasid vastava keele kõnesalvestistest jms.

- 🌸 Eesti murrete korpus
- 🌸 Automaatselt transkribeeritud raadiosaadete brauser
- 🌸 Eesti keele spontaanse kõne foneetiline korpus

Leksikaalsed ressursid

Sõnastikud, terminoloogilised ressursid, mõistete andmebaasid, sagedusloendid jms.

- 🌸 Eesti õigekeelsussõnaraamat ÕS 2013
- 🌸 Eesti etümoloogiasõnaraamat
- 🌸 Eesti-vene sõnaraamat
- 🌸 Kohanimede andmebaas

Tekstitötlusvahendid

Tekstiliste andmete töötlemiseks ja haldamiseks mõeldud tarkvara: speller, morfanalüsaator, sõnastike haldamise tarkvara, nimisõnafraaside märgendaja jms.

- 🌸 Eesti keele speller
- 🌸 Eesti Keele Instituudi morfoloogiatööriistad
- 🌸 Kollokatsioonide leidja veebiliides

Kõnetötlusvahendid

Kõneandmete töötlemiseks ja haldamiseks mõeldud tarkvara: tekst-kõne süntees, kõnetuvastustarkvara ning -rakendused jms.

- 🌸 Eestikeelne tekst-kõne süntees
- 🌸 Transkribeeriv diktofon Android operatsioonisüsteemiga telefonidele
- 🌸 Veebipõhine kõnetuvastus eestikeelsele kõnele

EKRK TEENUSED

Keeleressursside arhiveerimine ja haldamine

Tekstid ja korpused, keeleline audio- ja videomaterjal, sõnavarakogumid ja terminoloogiad, keeletötluse vahendid – kogu digitaalne keelevara on väärt talletamist! Tavaliselt võtab keskus teadaolevate keeleressursside valdajatega ise ühendust ning tutvustab neile oma võimalusi, kuid iga huviline võib ka ise pakkuda oma ressursse keskuses talletamiseks.

Ligipääs digiressurssidele

Arhiveeritud keelevarale on vaja ka virtuaalselt ligi pääseda. EKRK teeb kogutud digiressursid nii looja kui kasutaja jaoks turvaliselt ja mugavalt kättesaadavaks ühest kohast ja kokkulepitud tingimustel.

Kasutajate nõustamine ja koolitus

Ligipääs on ainult võimalus. Kuidas aga kogutud digiressursid kõige tõhusamalt enda jaoks tööle panna või kuidas leida digitaalsete keeleressursside abil vastuseid oma küsimustele – selleks pakub EKRK juhendeid ja KKK kodulehel, kuid teeb ka koolitusi ja vastab kasutajate päringutele.

Keskkond keeleressursside kasutamiseks ja arendamiseks

EKRK üheks ülesandeks on luua virtuaalne keskkond, milles asjatundlikud kasutajad saavad lisaks digivarast andmete kaevamisele ka keeleressurssi analüüsida, annoteerida ja uusi keeletehnoloogilisi vahendeid arendada.

EKRK KONSORTSIUM

EKRK tegutseb kolme partneri – TÜ Arvutiteaduse Instituudi, TTÜ Küberneetika Instituudi ja Eesti Keele Instituudi – konsortsiumina. EKRK juhtstruktuur ja keskeid teenuseid pakkuvad serverid asuvad TÜ ATI (www.cs.ut.ee) juures. TTÜ Kübl (www.ioc.ee) on juhtiv asutus kõnetehnoloogia alal, kõneressurssi töödeldakse sealses foneetika ja kõnetehnoloogia laboratooriumis. Sõnastikega seotud ressurssidega tegeletakse Eesti Keele Instituudis (www.eki.ee).

Uuri lähemalt:

keeleressursid.ee/et/keeleressursid