



Väike sissejuhatus keeletehnoloogiasse

Kais Allkivi-Metsoja

tlu.ee/~kais

Mis on teie jaoks keeletehnoloogia?

- Minge lehelele <https://www.menti.com>.
- Sisestage kood **5754 8469**.
- Tooge välja **kolm** märksõna, mis seostuvad keeletehnoloogiaga.



keele arendamine
tiktok roboti häälega
keeleõppimisprogrammid
eesti keele instituut
keelekasutus
helituvastus
grammatika kontrollija
suhtlemine läbi rääkimise
keeleõpperakendused
keeleõpe
masinõpe
teksti analüüs
tehiskratt
tõlge
kõnetuvastus
sorteerimine
hääletuvastus
lahe
sõnad
eki
ai
masintõlge
tehiskratt
andmestik
õs
keel
tõlkimine
keeletase
lingvist
duolingo
siri
korrektor
chatgpt
kõnes
grammatika
eesti keel
kõnesüntees
lingvist
duolingo
a word
duolingo öökull
ipä
arendamine
andmed
tekstitöötlus
õigekiri
statistiline analüüs
automaatkorrektuur
veebiotsing
grammarly
proge
tekst kõneks
veebisõnastik
keele töötlus
keeleõppimiskrakendused
another word
linguistics eesti keeles
llm-idega arvuti

Masintõlge

**Automaat-
korrektuur**

Vestlusrobotid

Veebiotsing

Kõnetuvastus

**Keeleõppe-
rakendused**

Kõnesüntees

„Kujutlus sellest, et arvutid võiksid osata kasutada inimkeelt, on sama vana kui kujutlus arvutitest endist.“

„Just arvutite võimekus käsitseda keelt samavõrd oskuslikult kui meie, inimesed, annab paljude jaoks märku masinate tõelisest intelligentsusest.“

(Jurafsky & Martin, 2009)

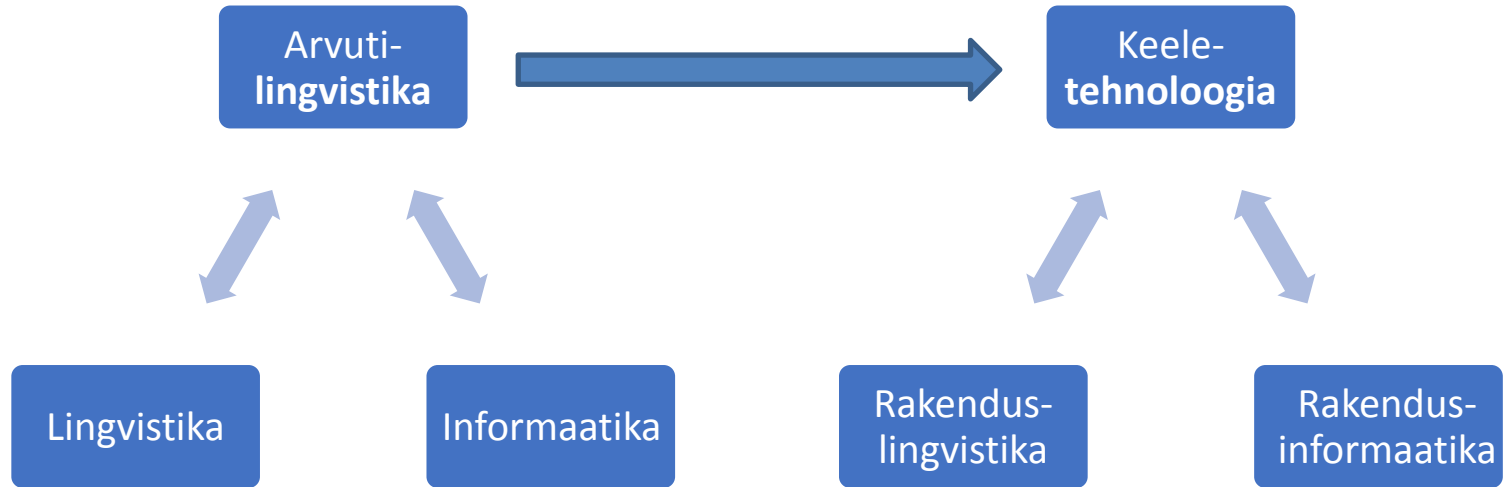


Keeletehnoloogia

- Keeleteaduse ja arvutiteaduse hübriideriala
- Tegeleb meetodite, tarkvara ja seadmetega, mis töötlevad tekste ja kõnet
- Sünonüüm: loomuliku keele töötlus (ingl *natural language processing – NLP*)



Arvutilingvistika rakenduslik pool



Eesmärk

- **Keeleteadmiste rakendamine** paremate arvutisüsteemide loomiseks:
 - inimese ja arvuti suhtluse parandamiseks;
 - info paremaks esitamiseks, kasutamiseks, otsimiseks ja analüüsimiseks;
 - inimkeele paremaks mõistmiseks ja töötlemiseks.

(Kaalep, 1999)

Keeleteadmiste tasandid

- Foneetika ja fonoloogia – häälikud ja nende kõla
- Morfoloogia – sõnavormide moodustamine
- Süntaks – kuidas sõnu omavahel ühendatakse
- Semantika – keeleüksuste tähendus
- Pragmaatika – kõneleja eesmärgid ja kavatsused
- Diskursiivne – lausungitevahelised tähenduseseosed

(Jurafsky & Martin, 2009)



Keeletehnoloogia harud keele vormi järgi

Tekstitehnoloogiad

- [õigekirja-](#) ja [grammatikakontroll](#)
- [nimetuvastus](#)
- [meelestatuse analüüs](#)
- sisu kokkuvõtmine
- teemade mudeldamine
- ...

Kõnetehnoloogiad

- [kõnetuvastus](#)
- [kõnesüntees](#)
- kõnelejatuvasus
- ...



Keeletehnoloogia harud kasutuseesmärgi järgi

Keeleanalüüs

- [süntaktiline analüüs](#) (sõnaliigid, sõnavormid ja lauseehitus)
- veatuvastus ja -parandus
- keeleoskuse hindamine
- nimetuvastus
- teemade ja meeletatuse analüüs

Keeleloome ehk -süntees

- [morfoloogiline süntees](#) (sõnavormi moodustamine)
- kõnesüntees
- sisu kokkuvõtmine
- dialoogsüsteemid
- masintõlge



Keeletehnoloogia meetodid

Masinõppesüsteemid

- algoritm õpib mustreid suurest näidistekstide kogudest
- sageli eri keeltele kohandatavad
- masin õpib kiiresti, aga kvaliteet võib oluliselt varieeruda

Reeglipõhised süsteemid

- keeleteadmised kodeeritakse reegliteks
- keelespetsiifilised
- täpsem kontroll keele töötluse üle, aga väga aja- ja töömahukas

Mõtteharjutus!

- Kuidas defineeriksid reegli „Mis on sõna?“
- Mille alusel saab määrata lausepiire?

Vt Pythoni näidet [siit](#).

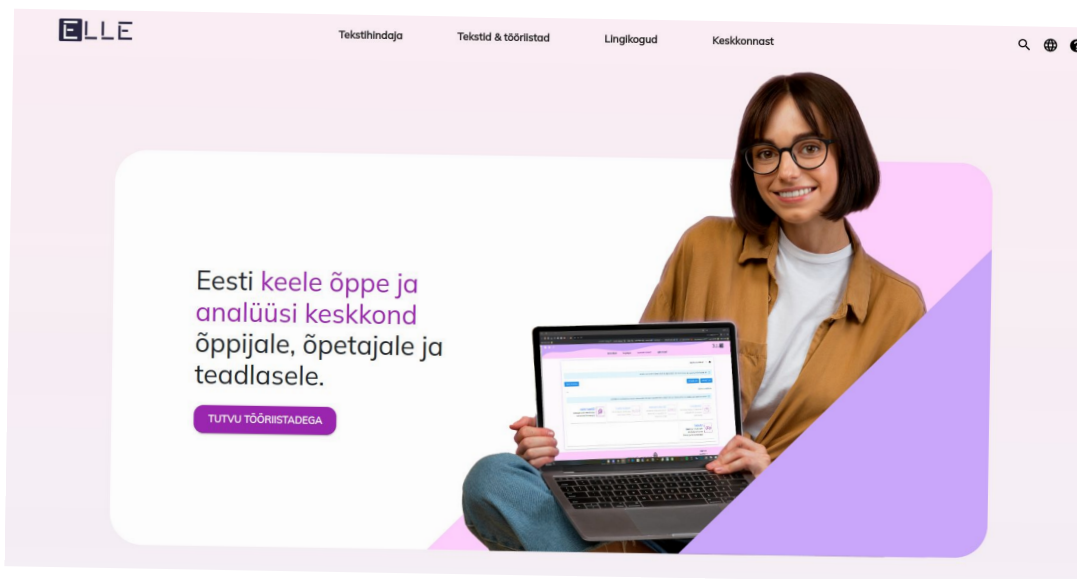
Keeletehnoloogia TLÜ-s

- Digitehnoloogiate instituudi üks teadus- ja arendussuundi
- Fookuses tekstianalüüs ja arvutipõhine keeleõpe
- Töörühm: <https://elle.tlu.ee/about/people>



Projektid (1)

- Keeleõppe ja -analüüsi keskkond ELLE (<https://elle.tlu.ee>)
- Sihtrühmi kaasav disain ja arendus
- Tekstikogud
- Tekstianalüüsi vahendid



Projektid (2)

- Eesti keele automaatkorrektuuri vahendite arendus: õigekirja- ja grammatikakontroll
- Koostöös Tartu Ülikooliga
- Kasutusnäide:

<https://colab.research.google.com/drive/1Gvy5aARFokGtLJ95VgIb6xyUB9RPFYsh?usp=sharing>



Projektid (3)

- ERR-i kultuurisaadete analüüs keeletöötuse abil:
 - nimetuvastus (isikud, kohad, organisatsioonid)
 - nimevahelised seosed (nt Arvo Pärt elas Berliinis)
 - teemad ja märksõnad
- Koostöös BFM-iga





Mida toob tulevik?

Keeletehnoloogia mõju suhtlusele

1. Tehnoloogia vahendab üha enam ja märkamatumalt meie vestlusi (kehal kantavad nutiseadmed)
2. Tehnoloogia on üha enam meie jutukaaslane (nt inimlaadsed vestlusrobotid tehistõelisuses)

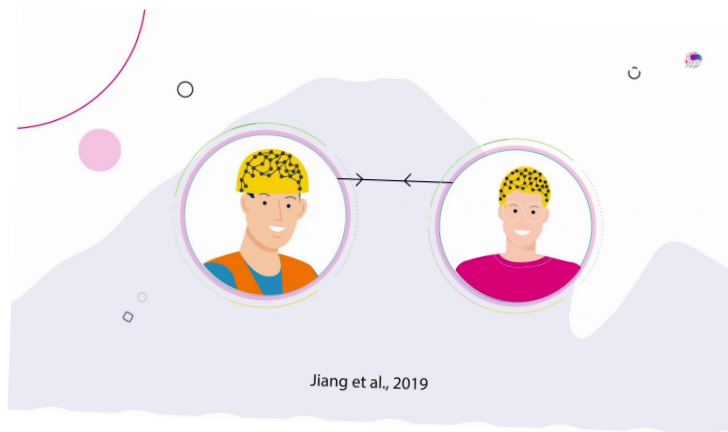


(Sayers et al., 2021)

Tulevikustsenaariumid

<https://lithme.eu/animations-and-survey/>

1. Masintõlgitud vestlus liitreaalsuses
2. Keeleõpe virtuaalreaalsuses
3. Viipe- ja väikekeelte mahajäämus
4. Kõneta suhtlus ehk nn mõtete lugemine



Viidatud allikad

- Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing*. 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall.
- Kaalep, H.-J. (1999). *Eesti keele ressursside loomine ja kasutamine keeletehnoloogilises arendustöös* [Dissertatsioon, Tartu Ülikool].
- Sayers, D. et al. (2021). *The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies*.

