

Näiteid keeletehnoloogia komponentide rakendamisest

Helen Kaljumäe, keeletehnoloog
keeletehnoloogia kompetentsikeskus | EKI



Riikliku keeletehnoloogia programmi projektid

HTMi koordineeritud riiklik programm keeletehnoloogia arendamiseks.

- „Eesti keeletehnoloogia 2018-2027“
- Hetkeseisuga 15 aktiivset projekti
- Neist mitu jätkuprojekti, nt kõnetuvastus, kõnesüntees, EstNLTK



Riikliku keeletehnoloogia programmi projektid

Kõnetuvastus

- Kõnetuvastus 2018-2022, jätkub, Taltech
- Inimkõne automaatne äratundmine ja selle tekstiks teisendamine

<https://github.com/taltechnlp/est-asr-pipeline>

EstNLTK

- EstNLTK teegi ja sellega seotud veebiteenuste arendamine 2018-2022, jätkub, TÜ
- Automaatsed analüüsivahendid eesti keele jaoks

<https://github.com/estnltk/estnltk>

ERR-i automaatsed reaalaajalised subtiitrid

- ERR, TalTech, EKI, HTM
- Kõnetuvastusel põhinev kiirkirjutaja lahendus subtiitrite automaatseks genereerimiseks.
- ERR-is saadaval reaalaajalised automaatsed subtiitrid.



Kõnetuvastus reaajas, subtiitrite loomine

- **TTÜ Kiirkirjutaja** <https://github.com/alumae/kiirkirjutaja>
 - sisendiks toores helisignaali
 - leia kõnega segmendid (ignoreeri vaikust, muusikat jne)
 - kas kõneleja on sama või muutunud?
 - tuvasta, kas eesti keel
 - tee kõnetuvastus
 - töötle loodud teksti (võõra sõna loomine, kirjavahemärgid, liitsõnade tuvastus, numbrid jne)
 - saada tulemus välja
- **Probleemid?**
 - Olenevalt materjalist vigu 5–20%
 - Eestikeelse kõne identifitseerimine
 - Aktsendiga kõne, sh murded, seeniorite ja laste kõne

Ärinimede kontrollija

- **Registrite ja Infosüsteemide Keskus (RIK), EKI**
- Ärinime valikul on mitmeid keelelisi nüansse, mida jälgivad nimekontrollijad ja muid seadusest tulenevaid nõudeid, mida kontrollivad kohtunikuabid. → püüame automatiseerida ärinime kontrollimise protsessi.
- Testperiood käimas.
- **Kasu automatiseeritud ärinimekontrollist:**
 - Kiire tagasiside valitud nime sobivuse osas
 - Väheneb RIKi registriosakonna koormus
 - Ettevõtte registreerimisprotsessi kiirendamine

Ärinimede kontrollija

Kontrollib kas...

... nimeväli on täidetud

... on lisatud tüübilaiend (OÜ, FIE, MTÜ jmt)

... on kirjapildilt või häälduselt sarnane mõnele olemasolevale nimele või kaubamärgile

... nimes esineb piiranguga sõnu (nt Eesti, linna, valla)

... nimes esineb musta või halli nimekirja sõnu

... nimes esinevad vaid lubatud tähemärgid (ei tohi nt [, %)

F - foneetiline sarnasus

T - tekstiline sarnasus

K - kaubamärgi kontroll

B - ebasobilike sõnade filter

L - kohanimede kontroll

G - riiklike terminite kasutuse kontroll

P - lubatud tähemärkide kasutuse kontroll

AITÄH

ktkk@eki.ee