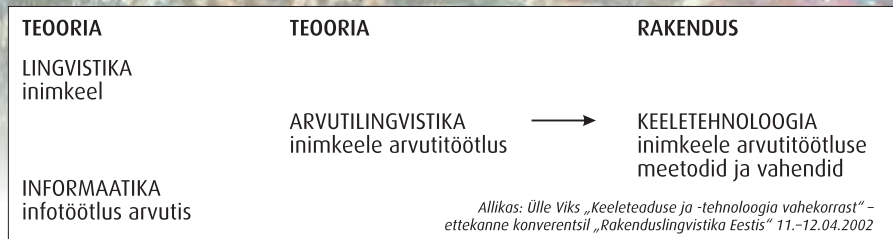


Eesti keeletehnoloogia

Mis on keeletehnoloogia ja arvutilingvistika?

Keeletehnoloogia on infomaatika haru, mis tegeleb inimkeele töötusega. Arvutilingvistika tähistab keeleteaduse ja arvutiteaduse hõõrieriala. Keeletehnoloogia ja arvutilingvistika tegelevad mõlemad loomuliku keele automaattöötusega, kuid arvutilingvistika läheneb probleemidele teoreetilisele ja keeletehnoloogia rakenduslikuma nurga alt.



Haldur Oim



Mare Koit



Ülle Viks

Missugused on arvutilingvistika ja keeletehnoloogia keskused?

Tartu Ülikoolis töötavad arvutilingvistika uurimisrühm ja keeletehnoloogia töörühm Eesti Keele Instituudis tegeldakse kõnesünteesi ja arvutileksikograafiaga Tallinna Tehnikaülikoolis on küberneetika instituudi foneetika ja kõnetehnoloogia labor

Õppida saab seda ala Tartu Ülikoolis, kus õpetatakse arvutilingvistika eriala filosoofiateaduskonna eesti- ja soome-ugri keeleteaduse õppekava raames ning keeletehnoloogia eriala matemaatika-informaatikateaduskonna infotehnoloogia õppekava raames. Kõnetehnoloogia-alaseid kursusi loetakse Tallinna Tehnikaülikoolis.

Millega tegelevad Eesti arvutilingvistika ja keeletehnoloogia?

Mõned näited:

Spelleri ehk automaatse õigekirjakorrektoriga puutub kokku iga arvuti abil teksti loov inimene. Õigekirjakorrektori aluseks on automaatne morfoloogiline analüsaator – programm, mis ütleb iga tekstisõna kohta, mis on selle sõna algvorm ja millises muutevormis on teda selles tekstis kasutatud. Nendele tekstisõnadele, mis jäävad morfoloogilise analüsaatori poolt ära tundmata, tõmbabki õigekirjakorrektor punase joone alla. Automaatset morfoloogilist analüüsi või selle pöördprotsessi – morfoloogilist sünteesi – läheb vaja ka infootsingul – näiteks selleks, et päringule *suveaeg* saaksime vastuseks ka lause, mis sisaldab väljendit *suveajale üleminek*.

Lisaks õigekirjakorrektorile tahaksid paljud kasutada ka grammatikakorrektorit, mis leiaks üles näiteks komavead kirjutatavas tekstis. Sellise kirjutaja abivahendi loomiseks on vaja süntaktilise analüüsi programmi, mis oskaks jagada lause osalauseteks ja tuvastada osalause te süntaktilise struktuuri.

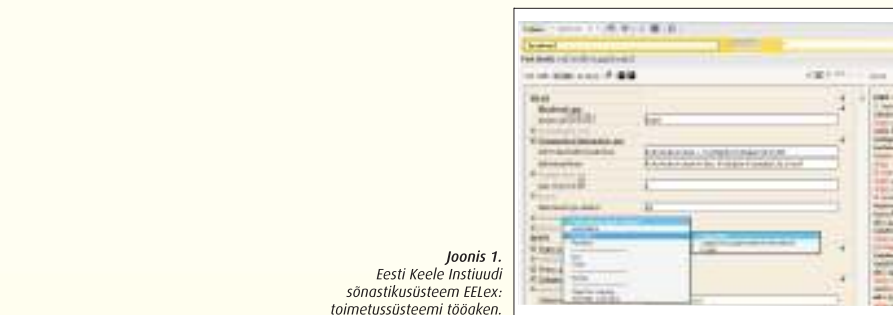
Sõnaraamatute koostajatel on suur abi veebipõhisest töökeskkonnast, mis integreerib erinevaid keeletehnoloogia vahendeid (keeletarkvara ja keeleressurss) selleks, et muuta sõnastike koostamine ja toimetamine lihtsamaks ja tõhusamaks.

Suur tulevik on dialoogsüsteemidel, st loomulikku keelt võimaldavatel kasutajaliidestel. Näiteks võiks infotelefonile vastata inimese asemel arvuti. Telefoni kaudu kasutatava dialoogsüsteemi loomiseks on vaja kõigepealt kõnetuvastust, mis teeb inimkõnest teksti, siis programmi, mis suudab inimese küsimuse „tõlkida“ andmebaasi päringuks ja päringu jälle inimesele vastuvõetavaks vastuseks. Selle tarvis on vaja teada, kuidas inimesed telefoni kaudu suhtlevad – küsimusi esitavad ja neile vastavad, ja lõpuks on vaja kõnesüntesaatorit, mis vastuse telefoni kaudu küsijale edasi annaks.

Siin on ajalehe Postimees veebiversiooni artikkel pealkirjaga „Kampaania käigus kogutud prügist suurem osa taaskasutatakse“, millest automaatne sisukokkuvõtete tegija hakkab kokkuvõtet tegema. Kasutaja saab valida kokkuvõtte mahtu: kas 10, 20, 30, 40 või 50% artiklist.



Automaatse sisukokkuvõtete tegija väljund: 30% kokkuvõtte artiklist „Kampaania käigus kogutud prügist suurem osa taaskasutatakse“.



Joonis 1.
Eesti Keele Instituudi sõnastikusüsteem EELex: toimetussüsteemi tööaken.

Dialoogsüsteem Teatriagent annab infot teatrite mängukavade kohta.



Joonis 2.
Eesti Keele Instituudi sõnastikusüsteem EELex: kasutajasisustiku loomise aken.



Veebiaadresse

arvutilingvistika uurimisrühm <http://www.cl.ut.ee>; keeletehnoloogia töörühm <http://www.cs.ut.ee/~koit/KT/>; Eesti Keele Instituut: <http://www.eki.ee>; TTÜ Kübi foneetika ja kõnetehnoloogia labor: <http://www.ioc.ee/>; Dialoogsüsteemid: <http://www.dialoogid.ee>
Tekst-kõne süntees <http://www.eki.ee/keeletehnoloogia/projektid/syntees/>;
<http://www.phon.ioc.ee/> -> projektid -> tekst-kõne süntees pimedatele
Elektronilisi sõnastikke koondavad portaalid <http://www.keeleeveeb.ee>;
<http://www.keelevara.ee>
Eesti Keele Instituudi sõnastikusüsteem EELex <http://exsa.eki.ee/>
automaatne sisukokkuvõtete tegija <http://math.ut.ee/~kaili/estsum/estsumframe.cgi>