Eesti keeletehnoloogia

Mis on keeletehnoloogia ja arvutilingvistika?
Keeletehnoloogia on infomaatika haru, mis tegeleb inimkeele töötlusega. **Arvutilingvistika** tähistab keeleteaduse ja arvutiteaduse hübriideriala. Keeletehnoloogia ja arvutilingvistika tegelevad mõlemad loomuliku keele automaattöötlusega, kuid arvutilingvistika läheneb probleemidele teoreetilisema ja keeletehnoloogia rakenduslikuma nurga alt.

Haldur Õim

Mare Koit

Ülle Viks

TEOORIA	TEOORIA		RAKENDUS
LINGVISTIKA inimkeel			
	ARVUTILINGVISTIKA inimkeele arvutitöötlus	→	KEELETEHNOLOOGIA inimkeele arvutitöötluse meetodid ja vahendid
INFORMAATIKA infotöötlus arvutis			

Allikas: Ülle Viks „Keeleteaduse ja -tehnoloogia vahekorrast" – ettekanne konverentsil „Rakenduslingvistika Eestis" 11.–12.04.2002

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:

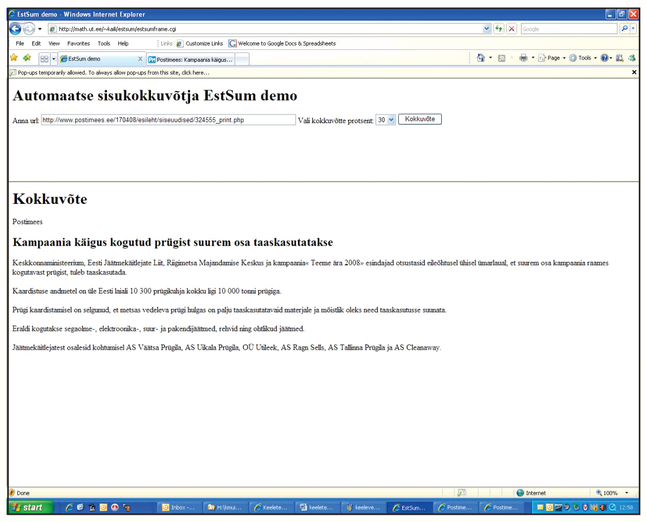
Spelleri ehk automaatse õigekirjakorrektoriga puutub kokku iga arvuti abil teksti loov inimene. Õigekirjakorrektori aluseks on automaatne morfoloogiline analüsaator – programm, mis ütleb iga tekstisõna kohta, mis on selle sõna algvorm ja millises muutevormis on teda selles tekstis kasutatud. Nendele 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 osalausete süntaktilise struktuuri.

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

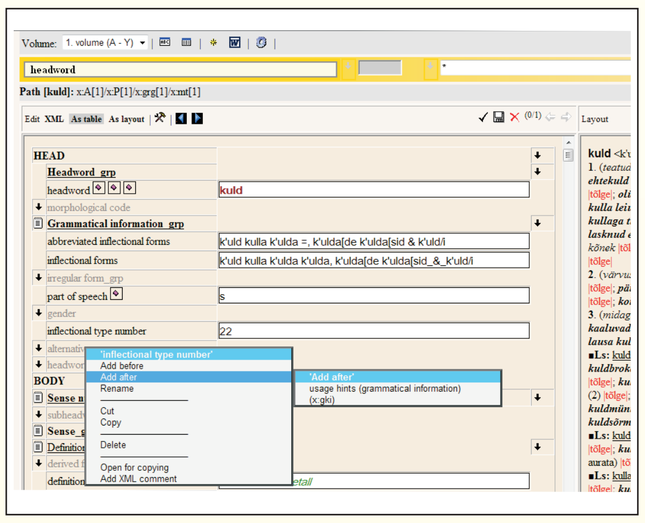
Suur tulevik on dialoogsüsteemidel, st loomulikku keelt võimaldavatel kasutajaliidestel. Näiteks võiks infotelefonile vastata inimese asemel arvuti. Telefoni kaudu kasutatava dialoogsüsteemi loomiseks on vaja kõigepealt kõnetuvastust, mis teeb inimkõnest teksti, siis programmi, mis suudab inimese küsimuse „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.

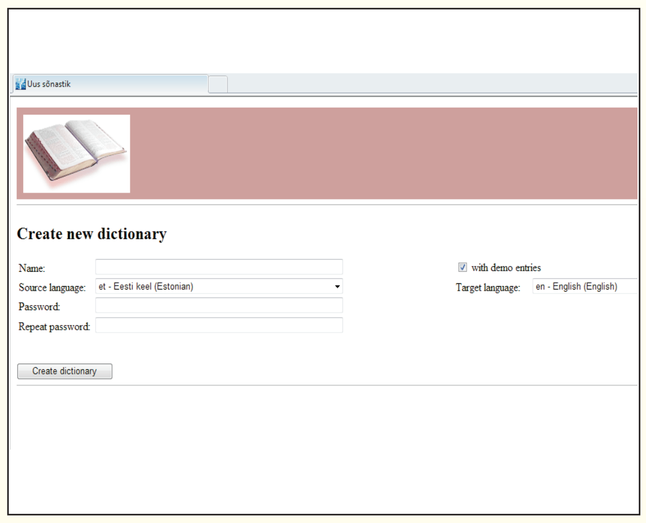


Automaatse sisukokkuvõtete tegija väljund: 30% kokkuvõte artiklist „Kampaania käigus kogutud prügist suurem osa taaskasutatakse".

Dialoogisüsteem Teatriagent annab infot Eesti teatrite mängukavade kohta.



Joonis 1. *Eesti Keele Instiuudi sõnastikusüsteem EELex: toimetussüsteemi tööaken.*



Joonis 2. *Eesti Keele Instiuudi sõnastikusüsteem EELex: kasutajsõnastiku 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
Elektroonilisi sõnastikke koondavad portaalid http://www.keeleveeb.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

EELex

Vaatmiku on teinud Emakeele Selts, Eesti Kirjandusmuuseum, Haridus- ja Teadusministeerium.
Tekst: Külli Habicht. Vormistus: Margus Nõmm, TÜ multimeediatalitus 2008.