

Kokkuvõte EKT aruandekonverentsi töögruppide aruteludest

Töögrupp 1: Millele riik peaks keeletehnoloogia toetamisel tähelepanu pöörama ja kuhu ressursse investeerima?

Temaatiliselt võib jagada töögrupi tulemid kolme valdkonda – eesmärgistamine keeletehnoloogia valdkonnas, õppurid ja õpe keeletehnoloogia valdkonnas ning EKT programmi kommunikatsioon ja selle korraldus.

Eesmärgid keeletehnoloogia valdkonnas

- Hetkel on ebaselge, millised on ootused keeletehnoloogia valdkonnas – kas oodatakse pigem rakendusi või rakendajaid? Ei ole õige väita, et inimesed valdkonnas ei sooviks midagi teha – häid mõtteid on tegelikkuses palju, kuid vaja oleks komplementaarsust. Keeletehnoloogia vajaks rohkem valdkonnasisest koostööd, mitte niiväga konkurentsi. **Ettepanek: mõelda sellele, kuidas suurendada valdkonnasisest sidusust ning võimaldada uutel huvilistel paremini keeletehnoloogiliste projektidega liituda.**

Keeletehnoloogia valdkonna õpe

- Keeletehnoloogia valdkonna õppetegevus ei ole hetkel riigi tasandil piisavalt toetatud. Selleks, et valdkond areneks ning oleks kasusaajate poolt enam vajalikuna tunnetatud, on tarvis keeletehnoloogilisi rakendusi. Samas rakendusi ei teki, sest valdkonnas ei ole 'kriitilist massi' tegijaid, noored ei ole alati seda valdkonda õppima ning valdkond kipub olema pigem fanaatikute päralt. **Ettepanek: keele- või kõnetehnoloogia valdkonnast võiks olla riiklik koolitustellimus.**
- Ülikoolid saavad hetkel ressursid, mille kasutamise üle otsustavad nad suures osas ise – riik ei saa palju suunata eriala moodustamist ega ka erialavalikut. Nt. Tartu Ülikoolis keeletehnoloogiat IT-erialana pole. Samas keeletehnoloogia õppekava vaid eestikeelsena ei pruugi leida turgu, ingliskeelsena poleks sellel aga väga mõtet.

Kommunikatsioon ja selle korraldus

- Riik soovib näha keeletehnoloogia valdkonnas enam lõpptarbija rakendusi ning leiab, et seda on võimalik saavutada läbi EKT programmi suurema nähtavuse ühiskonnas. Kommunikatsiooni teemasid hakatakse lahkama ka Eesti keele arengukava seire keeletehnoloogia töörühmas. Haridus- ja Teadusministeerium omalt poolt on valmis aitama keeletehnoloogia programmi kommunikatsiooni korralduses (nt. HTM-i kommunikatsioonibüroo kaudu). Esmaseks infokanaliks valdkonnas peaks olema keeletehnoloogia.ee kodulehekülg ja valdkonna e-posti loendid, kus tegevuse koordineerijaks on EKT programmi koordinaator. **Ettepanek: teha EKT programmi meediaplaan, mis käsitleks süstemaatilisemalt keeletehnoloogia valdkonna meediategevust.**
- Hetkel puudub hea ülevaade selle kohta, kes on erinevate projektide sihtgrupid ning võimalikud kasutajad, mida need sihtgrupid vajavad, et projektide tulemused oleksid kasutatavad/rakendatavad (enamasti kui lihtsalt prototüübid). **Ettepanek: viia läbi küsitlus projektijuhtide seas ning koguda selle kohta sisendit EKT projektide juhtidelt.**

Töögrupp 2: Millised on "valged laigud" ja "mustad augud" eesti keeletehnoloogias – täitmata eeldused, mis takistavad "Eesti keele arengukava" eesmärkide saavutamist?

Mustad augud ja valged laigud? Valgete laikudena saab tõlgendada erinevaid kaardistamata alasid ning valdkondi, millega keeletehnoloogia küll peaks tegelema, kuid millega ei tegeleta (erinevatel põhjustel – olgu selleks siis ressursside puudus, ekspertide puudus, avaliku huvi puudus vmt.). Mustade aukudena vaadeldi selliseid probleeme ja teemasid, kus millegagi küll tegeletakse, kuid ressurss läheb justkui kaotsi ja väljundit ei tule (või nt. jätkusuutlikkust ei teki).

Teises töögrupis käsitleti järgmisi teemasid: keeletehnoloogia valdkonna finantseerimine ja selle jätkusuutlikkusega seonduv, keeletehnoloogia valdkonna arenduste sisulised nõrkused ja puudujäägid, kommunikatsioon valdkonnas ning valdkondlik töökorraldus. Kõige tähtsam teema eeltoodutest on ehk finantseerimise temaatika, mis hõlmab endas mh. projektipõhise töövormi tavapärasest problemaatikast.

Keeletehnoloogia valdkonna finantseerimine ja rahastamise suunad

- Keeletehnoloogia valdkonna jätkusuutlikkus on suuresti seotud riikliku finantseerimisega. Kuna valdkond on seotud tihedalt eesti keeleruumiga, seega kasutajaskond piiritletud eesti keele rääkijatega, on valdkonnas keeruline pakkuda lisandväärtust, mis oleks sellest laiem. Sihtgrupi piiratusel tulenevalt lasub valdkonna arendamise toetamine suuresti riigi õul – nii nagu riik valdkonda finantseerib, nii jätkusuutlik see on. Valdkonna rakenduste sihtgrupp ei ole enamasti valmis arenduste eest (kasutamise eest) maksma (või kinni maksma kogu arendust). Keeletehnoloogia valdkonnas on suur mure selles, et rahastatakse küll projektide/keskkondade/rakenduste loomist, kuid kui projekt on läbi ja midagi on valmis tehtud, siis puudub finantseering/toetus projekti tulemuste ülalpidamiseks ja arendamiseks,

mistõttu palju sellest, mis tehtud, aegub, minetab tähtsuse ning muutub kasutuskõlbmatuks. Antud probleem on aktuaalne isegi selliste projekti tulemite puhul, kus kasutajaid väljaspool kitsast spetsialistide ringi oleks küll, kuid pole nt. ressursse, et ülal pidada juhendajaid, kes juhendaksid potentsiaalseid kasutajaid. Ettepanek on olnud seada sisse projektidele ka nõue, et juba projekti taotlemise hetkel annaks projekt märku, kuidas tagatakse projekti jätkuvus ja jätkusuutlikkus pärast projekti lõppu, kuid jääb küsimus – kas on õigus eeldada, et projekte arendatakse edasi ja hallatakse ilma igasuguste ressurssideta? **Ettepanek: luua meede, mis finantseeriks ka juba loodud projektide haldamist ja edasiarendamist ning tõhusamat sidumist uute projektide ja arendustega. EKT juhtkomitee võiks teha omapoolsed ettepanekud valdkonnaga seotud ministriumitele.**

- Mure on ka ettevõtlussektori ja selle rahastuse kaasamisega keeletehnoloogia valdkonda. Sageli on ülikoolidel probleem leida partnereid ettevõtlussektorist, kes oleks valmis projekte rahastama ning saaks projektide tulemil midagi ka rahastuse vastu. See, et tegemist on eesti keeletehnoloogiaga, mis on spetsiifiline ja väikese potentsiaalse kasutajaskonnaga valdkond, võimendab probleemi veelgi enam. Nt. chatbots – saaks sarnaseid lahendusi kasutada selleks, et veebilehtede külastatavus kasvaks – samas on selliste asjade loomine tasuvam pigem suurte keelte puhul, eesti keele puhul küsitav. Äriettevõtete võimalus teenida eesti keele pealt ei ole piisav, et investeerimine oleks mõttekas. **Ettepanek: ehk võiks olla riigi infosüsteemid see instants ja vahendaja, kes püüab ettevõtteid ja keeletehnoloogia projektide tulemeid lähendada. Nt. saadakse Keeleressursside keskusest teada, et on lõppenud konkreetsed EKT projektid selliste ja selliste tulemustega – kas nähakse perspektiivi edasiarendusteks?**
- Projektide finantseerimises puudub justkui mingisugune vaheetapp, mis oleks vahelüliks teadlase ja lõppkasutaja vahel. Näitena toodi sisukokkuvõtja – teadlane, kes laob vundamendi sellise rakenduse jaoks, ei tee valmis toodet. Tal puudub sageli selleks nii ressurss, aeg kui motivatsioon. Teadlane näitab läbi oma projekti ära, et see on võimalik, kuid tema töö tulemusel ei teki töötav ja laialtkasutatav programm – kuidas see peaks tekkima? Selliste rakenduste loomiseks oleks vaja oluliselt teistsuguse profiiliga projektijuhti (või ehk ka ettevõtjat), kes võtaks üle projekti tulemuse ja liiguks sellega edasi 'turukõlbliku' rakenduse poole. **Ettepanek: luua täiendav rahastusmeede (või mõelda olemasoleva meetmete fookuseerimise peale), mis keskenduks olemasolevate projektide tulemuste viimisele lõppkasutajate 'turule'. Järgnevatel programmi rahastamisvoorudes võiks mh. rahastada projekte, mis keskenduvad varasemate projektide tulemuste viimisele lõppkasutajani.**
- Ehk lahendaks järjepidevuse ja jätkusuutlikkuse see, kui riik ostaks mingite keeletehnoloogia valdkonna projektide puhul toote asemel teenust? Seega oleks tellimus nt. pigem see, et 'riik soovib, et programm toimiks ja töötaks', mitte, et 'mingiks ajaks on olemas selline toode'. See annaks tellijale ka garantii, et toimub pidev teenuse haldus, arendus ja parendus. Aitaks leevendada projektipõhisusest tulenevaid probleeme. **Ettepanek: riik võiks kaaluda teatud juhtudel keeletehnoloogia valdkonnas toote ostmise asemel teenuse ostmist.**

Keeletehnoloogia 'mustad augud' ning 'valged laigud'

- Kui vaadata META-NETi keeleraporti tabelid eesti keele kohta, on näha, et tehnoloogiline kaetus ehk ressursside olemasolu läheb langevas joones vastavalt keeletöötuse taseme tõusule: morfoloogiaressursside kaetus on hea, sellele tugineva süntaksi kaetus sobival tasemel. Keele genereerimise suunal on ressursse vähem kui analüüsimisel: morfoloogia on mingil määral olemas, süntaksit pole üldse, semantika samuti täielik valge laik. Põhjuseks võib tuua, et kõrgema taseme, näiteks semantika genereerimine on teaduslikus mõttes keeruline, samuti pole sellele rakendust. Samas, kui tulemus olemas oleks, siis leitaks ka rakendust, näiteks infootsing, kui tahetakse otsida eri allikatest infot ja genereerida saadud infost teksti. Mis tingimustel oleks võimalik säärane valge laik täita? Kuna tegu on küllaltki spetsiifilise projektiga, siis pole loota, et see rahastust saaks, kui tegu pole just julgeoleku suunaga, näiteks iisraeli pommirünnakute uurimisega. Säärased kõrgema taseme ressursside loomise katsed on kulukad ja neil pole garanteeritud väljundit, seega ei nähta niipea lootust nende loomise projektide rahastamiseks.
- Puudus on valminud ressursside edasiarendamise toest, seda nii kasutajani viimisel kui aegajalt kaasajastamisel, et muutuva tehnoloogiaga sammu pidada. Kui on arendusel keeletehnoloogiline programm, näiteks süntaksianalüsaator, antakse see projekti tulemusel deponooriumi üle. Kui aga mingil hetkel tuleks see kohandada 32-bitiselt arhitektuurilt 64-bitisele, pole see enam algse arendaja ülesanne. Samuti esitatakse programm sageli kujul, mis sobib küll muudes tööriistades kasutamiseks, kuid mitte lõppkasutajale tarbimiseks. Lõppkasutajarakendused, millesse ressursi integreerida, varieeruvad rohkem, kui ressurss ise, ja nõuavad vastavat kohandamist. Kui on vajadus mõningase arendamise-kohaldamise järele, enne kui olemasolevat ressursi saaks mõne IT firma rakendustes kasutada, siis firma ei tee mitte ettepanekut vahepealse osa arendamiseks, vaid jätab ressursi lihtsalt kasutamata. **Puudub asutus, mis tegeleks puuduolevate tükide väljatoomise, tellimise ja rahastamisega.**
- Eesti keeletehnoloogiliste ressursside kasutatavuse osas varieeruvad oluliselt ka erinevad tarkvaraplatvormid, näiteks kõnetehnoloogia pole Mac OS peal toetatud. Samas saab juba nt. Google Chrome'iga lasta suvalistelt internetilehtedelt teksti lugeda, samuti MS Wordis (kui on allalaetud hääl). Tore oleks, kui kõnesüntees ja -tuvastus hakkaks tööle Siri tüüpi süsteemidel, aga sinna vahele on semantikat vaja. Semantikata oleks kõnetehnoloogia kasutamise näiteks telefonirakendus „inimesed“ – saab otsida telefoniraamatust inimese, telefon loeb tema andmed häällega ette.
- Kas tuleks mõelda väljapoole traditsioonilist mõtlemist rakendusi oma ressurssidele, ka vahetulemustele? Näiteks süntaktilist ja semantilist analüüsi võiks kasutada masintõlke tulemuste hindamisel ja parandamisel. Inglise keelest eesti keelde tõlkimisel saadud lause keelise sobimatuse hindamine olemasoleva süntaktilise analüsaatoriga on mõeldav, kuid pole oodata kuigi head tulemust. Semantilise analüüsi kasutamine saadud lause mõttekuse hindamiseks on veelgi kahtlasem.
- Kerkis küsimus, kas Eestis ontoloogiatega tegeldakse? Ülemaailmne probleem on muuta dokumendipõhine internet andmepõhiseks, saada sealt teadmised kätte, selleks on aga oluline ontoloogiatega olemasolu. Ontoloogiatega eesti keelele tegeleb kümnekond inimest, on ka loodud kolm suurt ontoloogiat, mille abil püüti SEHKE süsteemis päringuid annoteerida. Peale seda, kui raha otsa sai ja projekt ära lõppes, on ontoloogiad jäänud õhku rippuma.

Kommunikatsioon ja valdkondlik töökorraldus

- Grupis toimus arutelu ka selle üle, kuidas on nähtav see, mida keeletehnoloogia valdkonnas tehakse ning kuidas meie potentsiaalsed kasutajad meid üldse leida võiksid. Näiteks toodi süntaksianalüüs, mida 'avalikkuse silmis' poleks justkui olemaski, seda ka teiste valdkonna teadlaste vaatepunktist (on olemas kaks valdkonna projekti, mis on vabalt kättesaadavad, kuid mida sellele vaatamata ei leita). Keeletehnoloogia kodulehel on üleval materjalid ja projektide tulemustest räägitakse, kuid kasutajad siiski projektide tulemite rakendamiseni ei jõua – mida valesti tehakse? **Ettepanek: peaks tegelema ka sellega, et projektide tulemid oleks paremini 'nähtavad' ja leitavad läbi otsingumootorite internetis.**
- Tuntakse, et lõppkasutajate hulgas keeletehnoloogiliste rakenduste vastu justkui huvi tuntakse, kuid see huvi ei jõua selleni, et tekiks konkreetseid rakendused ja potentsiaalsed kasutajad astuksid reaalseid samme rakenduste saamiseks/tellimiseks. Projektijuhid tunnevad, et ei osata piisavalt tõhusalt oma projektide tulemusi avalikkusele (atraktiivselt) tutvustada – nt. kes ja kuidas peaks tutvustama laiemale avalikkusele kõnesünteesi rakendusi või milline oleks see efektiivne kanal, mille kaudu seda teha? Uudiseid selle kohta on meedias ilmunud piisavalt, vahetult pärast uudise ilmutumist kasutajate hulk kasvab, kuid peatselt pärast seda vajub huvi ära. **Ettepanek: ehk peaks EKRR aktiivsemalt tegelema meediale ja ettevõtetele ressursside tutvustamisega? Sh. astudes ettevõtetega otse kontakti ja 'müües' projektide tulemusi. Samuti võiks teha sihtgruppidele ja potentsiaalsetele kasutajatele koolitusi, et tõsta teadlikust võimalikest rakendustest ja nende kasutamisest.**
- Keeletehnoloogia teenused peaks kindlasti toimima X-tee peal. Sellega peaksid tegelema ettevõtted, kuid kindlasti peaks juba tehtut pakkuma ka riik. Riigil on sageli ka suurem vajadus keeletehnoloogia järele kui ettevõtetel (sh. seetõttu, et toimub tihe suhtlus riikide vahel). Nii nt. töötatakse läbi suurt massi keelega seonduvat: tõlgitakse standardeid jmt. 80% vastavate inimeste ajast kulub tõlkimisele (sh. dokumendi koostamiseks, lugemiseks). Tõlkehangete maht kolmeks aastaks on suurusjärgus 300 000€, millest tekivad ka tõlkepaaride failid. **Ettepanek: panna riiklikesse hanketingimusse, et tõlkemälud antaks tulemusega kaasa (TLÜ magistr töö tõlkemälude autoriõigustest – peaks olema võimalik teha lepinguid nii, et tõlkemälud antakse tellijale üle. Rahvusvaheline praktika - kakskeelne tekst ja mälu kuulub tellijale).**
- Tegeleda oleks vaja selliste keeletehnoloogia valdkonna teemadega, mis lähevad organisatsioonide/asutuste üleseks. Nt. kui projektide tulemeid saab üles panna X-tee peale, siis kes neid tegelikult sinna üles paneb? Või hoolitseb hiljem selle eest, et need rakendused ka töötaksid? Projekti tulemuste ja rakenduste laiemaks levitamiseks on huvi olemas, kuid puudub koordinatsioonisüsteem. **Ettepanek: luua võimalused selliseks koostööks asutuste üleselt, et projektidel oleks võimalik hinnata, kes on projekti tulemitest ja rakendustest huvitatud (tutvustatakse projektide tulemusi otse huvitujatele), koostöö organisatsioonidele institutsioonidega oleks koordineeritud.**

Töögrupp 3: Millised peaksid olema EKT programmi tellitavate arendustööde prioriteedid?

Kolmandas arutelugrupis keskenduti eeskätt sellele, et luua mingisugune ideepagas erinevatest võimalikest arendustööde tellimustest, mida võiks EKT programmi raames rahastada. Peale selleteemalist ajurünnakut arutati ka selle üle, millised välja käidud ideedest võiksid olla kõige huvitavamad/prioriteetsemad. Lähteülesandena määratleti, et tellitavaid arendustöid peaks otsima eeskätt kolmest erinevast valdkonnast:

- Nn. 'augutäiteid' – sellised arendused ja tulemused, mis 'loomulikult teel' ei teki, kuid mis on puudu, mida oleks hädasti tarvis ning mis on valdkonnas takistuseks edasi liikumisel ja arenemisel;
- Väärtuslikud ja atraktiivsed 'lõppkasutajarakendused' – projektid valdkondades, kus on turunõudlus olemas, kuid pakkumise pool on nõrk – tellimuse esitamiseks võiks stimuleerida ka pakkumist;
- Selliste tulemuste ja projektide stimuleerimine, kus ollakse 'poolel teel' – nt. lõppenud projektid, mille tulemused on paljutõotavad ja huvitavad, kuid kus projektiga pole jõutud toimiva ja kasutuskõlbliku rakenduseni. Nõ. 'poolprototüüpidest' reaalsete ja töötavate lõppkasutajarakenduste väljaarendamine.

Alljärgnevalt on toodud loetelu töögrupi käigus loodud ideedepagasist (eespool on toodud ja ära märgitud ideed, mida peeti kõige atraktiivsemaks):

- Väga huvitav ja oluline: grammatikakorrektor LibreOffice'ile (ehk ka MS-ile);
- Väga huvitav ja oluline: kõnesünteesi ja kõnetuvastust kombineeriv lahendus, mis suudaks ette lugeda juhiseid (nt. korrata toidutegijale retsepti või auto parandajale anda juhiseid remondiks);
- Samuti huvitav: stiilivigade leidja;
- Samuti huvitav: vabateemaline küsimustele vastaja (midagi Siri sarnast? ... samas vähemate funktsioonidega – pigem wikist info otsija ... IBM'i Watsoni taoline?);

Muud ideed:

- Töötav parser (ei ole küll lõppkasutajalahendus);
- Praktiline kõnetuvastus, mis töötaks (nt. telefonimüüjad suudaksid oma kõned transkribeerida);
- Arvuti kasutajaliides kõnepõhiseks (nt. ava fail vmt. ... häälkäskluse põhine);
- Uuemad nutikad seadmed (nt. televiisorid) räägiksid eesti keelt (kõnetuvastus);
- Offline kõnesüntesaator Android platvormil (et saaks kasutada suvalises rakenduses - nt. uudiste lugemiseks);
- Keeleõpperakendused (eesti keele õppimine võiks olla lihtne - grammatikaharjutused jne.), täiendavalt ehk eesti keele häälduse õpetamine;
- Tekstitötlusrakendused – morfoloogiline analüüs teksti sees? (näidata sõnastikus, mis see sõna on jmt. lemmatiseerimine);
- Tekstianalüüsi veebirakendused/-teenused (pigem rakenduse tegijatele, mitte lõppkasutajatele) - nt. saad programmi saata mingi lause ja lausele tehakse semantiline analüüs;

- Meediamonitooringu firmad tahaksid sentiment analüüsi (positiivselt/negatiivne) ... peaks aru saada ka sellest, kellest negatiivselt/positiivselt räägitakse;
- Subjekt-verb-objekt võimalus analüütika instrumendina (teksti analüütika vahendina);
- Kopeeri/kleebi funktsioon, kus 'kleepimisel' oleks sõna õiges vormis.