

KEELETEHNOLOOGIAST TEHISINTELLEKTIKONVERENTSIDEL

2014. aastal oli meil võimalus käia kolmel tehisintellektikonverentsil, mis meie arust sobivad illustreerima viimaste aastate arenguid keeletehnoloogia valdkonnas, millesse meie endi töö on kuulunud. Seetõttu esitamegi muljed neist ühtse ülevaatenähtena.

Kolm konverentsi olid järgmised (ajalises järjekorras): 1) **ICAART-2014** (*International Conference on Agents and Artificial Intelligence*) toimus 6.–8. märtsil 2014 Lääne-Prantsusmaal Angers's (www.icaart.org/?y=2014); 2) **INISTA-2014** (*IEEE International Symposium on Innovations in Intelligent Systems and Applications*) toimus 23.–25. juunil 2014 Lõuna-Itaalias Alberobellos (www.inista.org) ja 3) **KEOD-2014** (*International Conference on Knowledge Engineering and Ontology Developments*), mis oli osa kolmikkonverentsist IC3K-2014 (vt allpool) ja toimus 21.–24. oktoobril Roomas (www.IC3K.org/?y=2014). Kahel esimesel osales Mare Koit, kolmandal Haldur Õim.

Põhjus neist koos kirjutada on eeskätt see, et ehkki tegu on rakendustele orienteeritud konverentsidega ja neist ühegi nimetuses ei esine mõiste *keel*, on ometi kõigi nende üks keskne probleem keele ja keelekasutuse (suhtluse) ning inimteadmuse ja -mõtlemise sidumine. Ühenduslülil on seega inimkeel, kas tekstidena, milles vajalikud andmed-teadmised on salvestatud, või kui vahend, mille abil mingi ülesande lahendamisel koostööd kavandatakse. Sellest vaatenurgast me neid konverentse tutvustamegi. Lisaks tasub märkida, et kõik kolm on n-ö sarikonverentsid, st on toimunud sama nimetuse all juba aastaid; see on andnud võimaluse jälgida, mida uut on vahepealse aja jooksul toimunud (meiegi oleme osalenud mitmel varasemal konverentsil).

Järgnevas anname konverentsidest ülevaate nende toimumisele vastupidises järjekorras, sest just ajalisel viimase konverentsi temaatika oli kõige laiem, kattes ühtlasi teiste konverentside temaatikat ja pakkudes seetõttu hea raami ka nendel konverentsidel käsitletud probleemide liigitamiseks.

KEOD ja IC3K

Nagu öeldud, on KEOD omakorda üks kolmest konverentsist, mis alati on toimunud koos, ühisnimetusega IC3K (*International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*); vaadeldav IC3K oli kuues. Teised kaks konverentsi on KDIR (*International Conference on Knowledge Discovery and Information Retrieval*) ja KMIS (*International Conference on Knowledge Management and Information Sharing*). Siduvaks mõisteks on seega *knowledge* (sellest ka 3K akronüümis) – otse tõlgituna 'teadmus, teadmised', aga tegelikult mõeldud katma pigem inimeste mentaalsfääri kogu sisu, ka nt arvamusi, seisukohti, oskusi, hinnanguid, emotsioone jne. IC3K peasponsoriks (niisamuti kui allpool tutvustataval konverentsil ICAART) on INSTICC (*Institute for Systems and Technologies of Information, Control and Communication*, rahvusvaheline teaduskonsortsium, mille peakorter asub Portugalis Setubalis). Kolm konverentsi on sel määral integreeritud, et olles registreerunud ühele neist, võib käia kuulamas ka teiste konverentside ettekandeid. Seda võimalust sai kasutatud ja järgnevas ei tutvustata ainult konverentsi KEOD ettekandeid, kus oli registreeritud meie ühisettekannet, vaid huvi pakkunud ettekandeid ka kaaskonverentsidelt. Kolmel konverentsil kokku esitati 275 ette-

kannet (neist KEOD 76) ning autoreid oli 30 riigist.

Ülevaates kommenteerime ettekanded, mis mingist aspektist seostuvad keelega, keeletehnoloogiaga, keele mõistmise modelleerimisega. Oleme siin vaadeldavad ettekanded jaganud kolme tinglikku temaatilisse rühma. Esimeses rühmas on ettekanded, kus kirjeldatud tööd ja tulemused jäävad peaaegu ainult keeletöötuse raamidesse: nt tekstidest sisukokkuvõtete tegemine, teatud valdkonna ontoloogia esitava semantilise andmebaasi koostamine, aga ka tekstide märgendamine neis kajastuvate hinnangute, suhtumiste ja tundmuste järgi (ingl *sentiment analysis*). Teises rühmas on tööd, mille sisuks ei ole lihtsalt tekstide, keelematerjali töötlemine uute keeleressursside loomiseks, vaid tulemuste kasutamine teatud suhtlusesmärkidel. Seega lisandub omaette probleemina vastavate suhtlussituatsioonide modelleerimine. Siia kuuluvad nt nõustamisüsteemid, otsuste tegemise tugisüsteemid, teatud suhtlustüüpe (nõupidamised, läbirääkimised, debatid) modelleerivad süsteemid. Kolmandasse rühma on paigutatud tööd, mille puhul teksti töötlemise fookuses on kindlat tüüpi kasutaja või kasutajate kategooria – nii selles mõttes, et püütakse kaardistada konkreetse kasutaja spetsiifilisi vajadusi, huvisid, eelistusi (kasutaja profiil), kui ka selles mõttes, et kui kasutaja profiil on teada, siis suudetakse olemasolevate andmete (tekstide) hulgast välja valida need, mis just sellele kasutajale võivad huvi pakuda (kusjuures andmed võivad olla nii äritekstit, turismiinfo, keeleõppematerjalid, uudised, ilukirjandus, kuni muusikapaladeni välja, ning kasutaja võib olla nii mingi kollektiiv kui üksikisik – nt keeleõppes). Rühmad pole muidugi selgete piiridega ega sõltumatud: nt esimeses rühmas kirjeldatavaid töötlusvahendeid läheb vaja nii teise kui ka kolmanda rühma töödes, kolmanda rühma töid (süsteeme) võib enamasti vaadelda kui teatud nõustamis- või otsuseid toetavaid süstee-

me, ainult et need on kasutajakesksed, mitte teema-, probleemi- või suhtlustüübikesksed.

Esmalt lühidalt kolmele IC3K konverentsile ühistest plenaarettekannetest (ingl *keynote lectures*). Kommenteerime neid vaid valikuliselt, sest keeletöötuse või keelelise suhtlusega need vahetult ei seostu, küll aga võib juba pealkirjadest välja lugeda konverentside üldise temaatilise raamistiku, mille sisse keeletöötus- ja suhtlusanalüüsi probleemistik konverentsidel paigutus. Plenaarettekandeid saab vaadata-kuulata aadressil www.ic3k.org/PreviousInvitedSpeakers.aspx#2014.

Domenico Talia (Itaalia) käsitles suurte andmete (ingl *big data*) töötlemist. Suurte andmete all ei mõelda lihtsalt mahult suuri andmekogumeid, vaid sellist andmekogu, mis ka kasvab ja muutub väga kiiresti ning lisaks sellele on heterogeenne, st andmed pärinevad erinevat tüüpi allikatest, on erineva struktuuriga jne. Suurte andmete töötlemiseks on vaja erilisi meetodeid, tavapärased andmetöötuse meetodid selleks ei sobi.

Sonia Bergamaschi (Itaalia) ettekandne oli samuti pühendatud suurtele andmetele: kuidas selliste andmekogudega töötada (neid struktureerida, märgendada, neist mingile päringule vastavat infot leida). Mõiste, mis kordus ettekandes erinevates seostes, oli *intelligentne* (ingl *intelligent information integration, intelligent search strategies*), osutusega vajadusele kasutada tehisintellekti vahendeid.

Wil van der Aalst (Holland) rääkis protsessikaevest (ingl *process mining*). See on uus uurimisvaldkond võrreldes tavalise andmekaevega: ei keskenduta (staatilistele) faktidele, seisunditele, situatsioonidele, vaid ajas kulgevatele sündmustele kui sidusatele protsessidele, mis, ehkki jagatavad üksikaktideks, moodustavad pideva terviku (vrd film *versus* slaidide järjend). Protsessikaeve vahendite abil saab jälgida ja analüüsida nt äriprojekti (või mis tahes kollektiivse

tegevuse) kulgemist tervikuna ja ka iga üksiku osalise tegevusi selles – kui see muidugi on dokumenteeritud (vt www.processmining.org). Lähenemine pakub kahtlemata huvi ka keeleanalüüsis, nt semantikas ajas kulgevaid protsesse kirjeldavate tekstide (või suhtlusanalüüsis dialoogide) analüüsimisel, kui on vaja teksti eri kohtades kirjeldatud üksikakte, seisundeid (või dialoogis repliike) tuvastada mingi sündmuse või episoodi osadena.

Marie-Jeanne Lesot (Prantsusmaa) käsitles üleminekut teadmuse objektiivselt esituselt subjektiivsele tõlgendusele. Ettekanne sobib illustreerima üht viimaste aastate silmatorkavamaid tendentsi vaadeldavas valdkonnas: liikumist andmete semantilise/ontoloogilise analüüsi ja kontseptuaalse tõlgendamise juurest neis sisalduva emotsionaalse jm suhtumusliku info tuvastamise poole, nt kommentaariumide tekstid või Twitteri säutsud, aga visuaalsete andmete puhul (TV diskussioonisaated) ka keha-keel, näoilmed. Ning siin läheb vaja ka teistsuguseid andmetöötlusmeetodeid ja -vahendeid: hägus klasterdamine (ingl *fuzzy clustering*), afektiarvutus (ingl *affective computing*).

Nüüd konverentsiettekannetest eespool kirjeldatud rühmade kaupa.

1. teema: keeletöötlus

Ahmed Tawfiki jt (Itaalia) ettekanne mitmekeelsetest keeleressurssidest (KEOD-il) sobib illustreerima ka meil Eestis aktuaalse keeleressursside teema käsitlust. Kriitilised terminid on *mitmekeelsus* ja *interaktiivsus*, aga põhiteema on see, kuidas mitmekeelseid keeleressursse saaks kasutada vajaliku teadmuse hankimiseks infootsingul, sõltumatult päringu esitamise ja otsitavate tekstide keelest. Ehk siis keskne probleem on tekstide märgendamiseks kasutatavad mõistepõhised metaandmed ja nendega töötamine (kus omakorda on kriitilise tähtsusega väljendite automaatne tõlkimine). Tutvustati selliseks tööks loodud

platvorme. Teiseks näiteks sobib Nouha Mhimdi jt (Tuneesia) ettekanne tekstide sisukokkuvõtete semantiliseks anno-teerimiseks vajalike ontoloogiate konstrueerimisest (KEOD-il).

Järgmised kaks KDIR-i ettekannet on aga näide sellest, kuidas eespool mainitud huvikeskme nihkumine tekstide kontseptuaalselt sisult neis leiduvatele hinnangulistele aspektidele (lihtsatest arvamustest emotsioonideni) võib realiseeruda tekstide märgenduses ja analüüsis. Twitteri tekstide analüüsi käsitlevas ettekandes võrdlesid Evangelos Psomakelis jt (Kreeka) säutsude kui tekstitüübi erinevaid esitus- ja analüüsi-meetodeid, tuginedes enam kui 4000 kätsi märgendatud säutsu analüüsile, ja leidsid, et ülekaalukalt paremad on õpipõhised (ingl *learning-based*) meetodid. India teadlaste Prateek Nagwanchi jt ettekande teemaks oli aga konkreetset ühe emotsionaalse taustaga suhtumise – sarkasmi – tuvastamine tekstis. Analüüsiti nii lingvistiliste (leksikaalsemantiliste) kui ka statistiliste andmete kasutamise võimalusi ja leiti, et parimaid tulemusi annab nende kooskasutamine, rõhutades samuti õpipõhise lähenemise olulisust. Väärrib eraldi märkimist, et tundmusanalüüsis (ingl *sentiment analysis*) on just sarkasmi ja ironia kaks erilist tähelepanu pälvinud suhtumist, ju seetõttu, et need mitte lihtsalt ei varjunda teksti emotsionaalselt, vaid muudavad teksti mõtte sõnasõnalisega võrreldes vastupidiseks.

2. teema: suhtlus

KEOD-il kuulus sellesse valdkonda käesoleva ülevaate autorite stendiettekanne, mille esitas H. Õim ja mis tutvustas debati formaalset mudelit. Osalejatel on debatti alustades vastandlikud eesmärgid, mille kasuks nad mõlemad esitavad debati käigus argumente. Lõpptulemusena saavutab üks neist oma eesmärgi (võidab debati), aga teine peab oma eesmärgist loobuma (kaotab).

KMIS-il peetud ettekannetest toome

esile kahte. Volker Wagner (Saksamaa) rääkis sellest, kuidas modelleerida ühise ülesande lahendamisel koos toimivat meeskonda. Ta esitas nn ühismälu mudeli, mis hõlmab ühelt poolt ülesande jaotamise erinevateks osaülesanneteks, vastavalt meeskonna liikmete kompetentsile, ja teiselt poolt teadmuse jagamist osaülesannete lahendamisel. Elisabetta Fersini jt (Itaalia) ettekandes käsitleti arendatavat veebipõhist kohtuvaidluste lahendamise süsteemi eMediation. Süsteem sisaldab andmete kogumise keskkonda, mis võimaldab määrata kohtuvaidluse sisu, ja varasemate kohtuasjade infootsingu, mis aitab vaidluses osalejatel aru saada oma kohustustest. Süsteemi kui vahendaja ülesandeks on otsida ja rakendada optimaalset strateegiat vaidluse lahendamiseks.

KDIR-ilt mainime Youssef Meguebli jt (Prantsusmaa) ettekannet hoiakute (ingl *opinion*) väärtustamisest, tuginedes veebikeskkonnas peetud debattidele kasutajate vahel. Hoiakute arv, mida kasutajad väljendavad nn uues meedias, kasvab plahvatuslikult; inimesed kulutavad üha enam aega mitmesuguste teemade arutamiseks ja oma kogemuste jagamiseks. Sellise kasutajate loodud konteksti arvessevõtmine võimaldaks tõsta uudiseartiklite soovitamise ja otsingu efektiivsust. Hoiakute korpus on aga sageli hiigelsuur ja sisaldab palju müra ning see teeb olulise konteksti leidmise raskeks. Defineeriti hoiaku väärtus (olulisus), võttes arvesse tema seoseid teiste väljendatud hoiakutega. Selleks loodi hoiakute graaf, kus iga serv esitab tundmust (ingl *sentiment*), mida hoiak väljendab teise hoiaku suhtes, ja esitati algoritm, mis suurendab positiivsete ja vähendab negatiivsete tundmustega hoiakute väärtusi. Katsetes, kus kasutati andmeid kolmest veebiarhiivist (*CNN*, *Independent* ja *The Telegraph*), ilmnes, et selline hoiakute väärtustamise mudel annab häid tulemusi.

3. teema: kasutajaprofil

KEOD-il peetud ettekannetest märgime kahte. Artemis Parvizi jt (Ühendkuningriik) käsitlesid seoses semantilise veebiga loogiliste järelduste tuletamist formaalses ontoloogias. Kasutajale pole vaja esitada kõikvõimalikke järeldusi, vaid nende hulgast tuleb teha valik, vastavalt konkreetse kasutaja vajadustele. Ettekandes põhjendati, et olemasolevates süsteemides ega ka kirjanduses pole valikuprobleemile seni pööratud piisavat tähelepanu, ja pakuti mõned lahendused. Gharsa Ali jt (Egiptus) käsitlesid soovitusüsteeme, mille eesmärk on aidata kasutajal teha õige valik erinevate võimaluste hulgast, sõltuvalt tema eelistustest ja huvidest. Esitati semantikal põhinev soovitusüsteem, milles luuakse kasutajaprofil ja seotakse see valdkonnaontoloogiaga. Eksperimentides reaalsete andmetega tõestati, et ontoloogiaga seotud soovitusüsteem on parem kui senised, kus ontoloogiast ei kasutata.

KDIR-i ettekannetest toome esile kolm. Ghada Besbes jt (Tuneesia) tutvustasid personaliseeritud infootsingusüsteemi, mis põhineb hägusontoloogial (ingl *fuzzy ontology*). Hägusontoloogiad kasutavad nn hägusloogikat, võimaldades käsitleda ka ebatäpset infot ja seega paremini arvesse võtta kasutaja vajadusi. Süsteemi kuulub kasutajaprofili mudel, mis infootsingu efektiivsuse tõstmiseks arvestab sama kasutaja varasemaid otsinguid ja eelistusi. Servet Tasci jt (Türgi) esitlesid veebipõhist uudiste soovitamise süsteemi, mille koosseisu kuulub kasutajaprofil, mis võimaldab konkreetset kasutajat leida oma huvidele ja vajadustele vastava info. Süsteemis on ka uudiste veebist kogumise, sisupõhise liigitamise ja sisukokkuvõtete tegemise moodulid. Fumiyo Fukumoto jt (Jaapan) tutvustasid süsteemi, mis soovitatav külastajale tema vajadustele vastavat hotelli.

INISTA

Rahvusvahelist sümposiooni INISTA korraldab alates 2005. aastast elektri- ja elektroonikainseneride instituut IEEE (*Institute of Electrical and Electronics Engineers*). 2014. aasta sümposiooni toimetised on saadaval IEEE Xplore digitaalses raamatukogus (ieeexplore.ieee.org/xpl/conhome.jsp?punumber=1803839).

Osalejaid oli umbes 100, kuulati kokku 70 suulist ettekannet, millest esimene oli plenaartekanne. Üritus algas kolme õpikojaga, millest küll ükski ei olnud seotud loomuliku keele töötusega. Enamik siinses ülevaates käsitletavatest ettekannetest olid keele automaattööluse alased, st kuulusid eespool toodud liigituse kohaselt esimesse rühma.

Plenaartekande pidas Vicenco Loia (Itaalia) uuest, arvuti- ja sotsiaalteadusi ühendavast uurimissuunast – tundmusarvutusest (ingl *sentic computing*). Tundmusarvutuse eesmärk on tundmuste, hoiakute ja emotsioonide väljasõelumine loomuliku keele tekstidest, mis aitab teksti paremini mõista. Tundmusarvutust saab täiendada hägusloogikaga – lingvistilise ebakindluse modelleerimise vahendiga, mis võimaldab analüüsida ka emotsioone. Tundmusanalüüsi praktilise rakendusena tõi esineja välja sotsiaalse veebi: veebikasutajad kalduvad üha enam väljendama oma arvamusi ja hoiakuid igapäevaelu, veebiteenuste ja -toodete jms kohta. Samal ajal on e-teenuste (e-äri, e-turism jne) levikul oluline mõju kasutajate hinnangute ja otsuste kujunemisele ning neid hinnanguid arvestavad teised kasutajad sageli rohkem kui ekspertide seisukohti. Inimliku mõõtme toomine teksti automaatsesse analüüsi tagaks seda, et suudetakse ette ennustada ja arvesse võtta kasutajate tundmusi, ootusi ja eelistusi. Firmad saavad kasutada tundmusanalüüsi tulemusi oma turundusstrateegiates. Esineja juhtis tähelepanu leksikaalsele ressursile SentiWordNet (sentiwordnet.isti.cnr.it/), mis põhineb

Wordneti sünohulkadel ja võimaldab tekstidest automaatselt välja sõeluda hoiakuid/tundmusi.

Mitmes ettekandes vaadeldi emotsioonide automaatselt tuvastamist. Näiteks esitas Cigdem Eroglu Erdem (Türgi) meetodi emotsioonide tuvastamiseks audio-videosalvestustes, mis ekstraheerib tunnused nii kõnest kui ka videost ning seejärel kombineerib neid, arvestades tunnuste väärtuste haripunkte. Vitoantonio Bevilacqua (Itaalia) tutvustas programmi kuuesteistkümmene erineva nimžesti tuvastamiseks, mille alusel ennustatakse seitset erinevat emotsionaalset seisundit.

Kaks ettekannet Türgist käsitlesid türgikeelsete tekstide automaattöölust. Sahin Isik tutvustas türgikeelsete spämmimeilide identifitseerimist kolmel tuntud statistilisel meetodil; parimaks meetodiks osutus hii-ruut. Zeynep Banu Özger vaatles küsimuste tuvastamist türgikeelsetes Twitteri säutsudes, olles selleks eelnevalt kogunud ja märgendanud ühemiljonilise säutsude korpuse. Kõigepealt rakendati korpusele eeltöölust, et eemaldada korduvad säutsud. Siis leiti reeglite abil küsimuste kandidaadid ja lõpuks valiti sellest hulgast statistikat kasutades säutsud, mis kõige tõenäolisemalt võisid olla küsimused.

Vincenzo Di Lecce (Itaalia) käsitles itaaliakeelse teksti ühestamise süsteemi, mis loodi semantilise veebi tehnoloogiat kasutades, et tõhustada kasutajate suhtlust veebis.

M. Koit (Eesti) vaatles eestikeelsete argumenteerimisdialogide struktuuri automaatselt tuvastamist.

Ettekandeid peeti veel signaalitöölusest, robotikast (eeskätt mobiilsetest robotitest), meditsiinilise diagnoosimise süsteemidest, liikluse reguleerimise süsteemidest, tehisintellekti alustest jne.

Üks sessioon oli tervikuna pühendatud intelligentsete süsteemide loomisele Itaalia ühes regioonis, nn elulaborite programmile (Living Labs, livinglabs.regione.puglia.it/inista-2014), mis muu-

hulgas sisaldab keeletehnoloogia rakendamist ja konkreetse kasutaja vajaduste arvessevõtmist. Elulaborid on regionaalne uurimis- ja innovatsioonistrateegia aastateks 2014–2020, kus teadlased, ettevõtted ja elanikegrupid on aktiivselt kaasatud uute infotehnoloogiliste lahenduste arendamisse ja testimisse, vastavalt nende spetsiifilistele vajadustele. Ideid ja teadmisi vahetades loovad nad üheskoos „avatud ökosüsteemi”, milles lõppkasutajad osalevad uute toodete ja teenuste kavandamisel, rakendades selleks info- ja kommunikatsioonitehnoloogiat. Praeguseks on regioonis 79 aktiivset elulaborit mitmes erinevas valdkonnas, nagu elukeskkond, julgeolek, kultuuripärand ja turism, e-valitsus, tervishooldus. Mõnesid neist esitleti konverentsil.

ICAART

Agentide ja tehisintellekti konverentsi ICAART korraldati juba kuuendat korda. Kokku oli umbes 200 osavõtjat 48 riigist. Esitati kuus kutsutud ettekannet, 83 suulist ja 40 posterettekannet. Ilmus toimetiste kogumik, mis on kättesaadav andmebaasist *dblp* (www.informatik.uni-trier.de/~ley/db/conf/icaart/). Töö toimus paralleelselt kahes sektsioonis: tehisintellekt ja agendid. Konverents algas paneeliga, kus esinesid needsamad teadlased, kes hiljem pidasid plenaarettekanded. See andis hea ülevaate ettekannete temaatikast. Plenaarettekannetest kolm võib tinglikult paigutada eespool toodud liigituse kohaselt esimesse rühma, st need olid seotud (ka) keele mõistmisega.

Philippe Leray (Prantsusmaa) rääkis Bayesi võrkudest, mis võimaldavad graafiliselt esitada andmetes sisalduvaid teadmisi ja läbi viia arutlusi, kui andmed ei ole täielikud või on ebatäpsed. Loeng keskendus algoritmidele, mida kasutatakse Bayesi võrgu õppimiseks sõltuvalt andmete iseloomust ja eelneva teadmuse olemasolust.

Claude Frassoni (Kanada) ettekanne oli pühendatud emotsioonide mõõtmisele, mis on tähtsal kohal õppimisprotsessis. See tingib vajaduse füüsiliselt mõõta ajutegevust ja inimese vaimset aktiivsust selgitamiseks välja õppimisprotsessis avalduvaid negatiivseid ja positiivseid emotsioone. Selle põhjal saab omakorda anda pedagoogilisi soovitusi ja luua õppimiseks vajalikke emotsionaalseid keskkondi, et parandada õpitulemusi.

Jaap van den Herik (Holland) rääkis suurte andmete töötlemisest. Traditsioonilised andmetest teadmiste väljasõelamise meetodid eeldavad, et andmed on eelnevalt struktureeritud ja relevantne info on ette teada. Suurte andmete korral ei ole see eeldus täidetud. Teine erinevus on, et suurtest andmetest ei otsita põhjuslikke seoseid, vaid hoopis koosinemisi. Esineja tõi näite kuritegude ennetamisest: suure andmestiku alusel leiti, et soojade ilmadega pannakse supermarketite ümbruses toime rohkem kuritegusid. Muidugi ei saa väita, et ilma ja kuritegude arvu vahel eksisteeriks põhjuslik seos, kuid leitud korrelatsioon oli piisav praktiliseks järelduseks, et ilmade soojenedes tuleb supermarketite lähedusse suunata rohkem politseinikke. Seega on oluline suurtest andmetest leitud koosinemiste tõlgendamine. Kasutatakse visualiseerimist, mis aitab mõista andmetega kirjeldatavat „lugu”. Paljudes valdkondades on juba toimunud üleminek relatsioonidelt korrelatsioonidele, esineja nimetas seda arvutuslikuks pöördeks (ingl *computational turn*). Sedamööda, kuidas tõuseb arvutiprogrammide intelligentsus, luuakse uusi arvutusmudeleid, mis võimaldavad viia läbi arutlusi suurte andmetega ja tuletada korrelatsioonidest ka põhjuslikke seoseid andmete vahel.

Ülejäänud kolm plenaarettekannet olid (osaliselt) seotud keeletöötlemise tulemuste rakendamisega suhtluses, sh kindla kasutaja arvestamisega, st kuuluvad teise või kolmandasse rühma.

Henry Lieberman (USA) kõneles kasutajaliidetest. Kasutajaliidese kavandamisel on seni enamasti eeldatud, et kasutaja teeb oma valikuid väikesest operatsioonide hulgast, ja liides on üles ehitatud menüüvalikute või ikoonidena. Järjest enam ilmub aga selliseid rakendusi, kus operatsioonide hulk ulatub sadadesse ja tuhandettesse ning kasutaja suhtleb rakendusega loomulikus keeles teksti või kõne vahendusel, žestide abil jne. Sellisel juhul on otstarbekas kavandada kasutajaliideseid lähtuvalt kasutaja eesmärkidest, kus vajaminevad operatsioonid grupeeritakse alles töö käigus. Ettekanne tutvustas tehisintellekti strateegiaid selliste olukordade käsitlemiseks.

Matthias Kluschi (Saksamaa) ettekanandes käsitleti intelligentsete otsustusüsteemide (nt soovitusüsteemid, kokkuleppimissüsteemid) loomist, kasutades agenditehnoloogiat ja semantilist veebi.

Pieter Sproncki (Holland) ettekanne andis ülevaate nn virtuaalsete agentide kasutamisest mängudes, sh treenimiseks ja õpetamiseks.

Sektsioonides käsitlesid mitmed ettekanded keele mõistmise modelleerimist, st kuuluvad eespool toodud liigituse kohaselt esimesse rühma. Michael Harré (Austraalia) vaatles vaimuteooria (*Theory of Mind*) rakendamist. Inimesel on arenenud kognitiivne mehhanism, nn vaimuteooria, mis võimaldab meil mõista teise isiku psühholoogilist ruumi: tema motivatsiooni, kitsendusi, plaane, eesmärgi ja emotsionaalset seisundit. Tänu sellele võimele saame aru teiste isikute tehtavatest valikutest, sest neil on oma „sisemine maailm”, mis mõjutab nende valikuid samamoodi, nagu meie sisemine maailm mõjutab meie omi. See on ehk kõige tähelepanuväärsem kognitiivne areng inimese evolutsiooni ajaloos, samuti meie võime kohaneda tuttavate olukordadega, aga ka arutleda dünaamiliselt täiesti uues olukorras. Nii tekib küsimus: kas me võime rakendada

inimese kognitiivsuse selliseid aluseid nagu vaimuteooria ka tehisevaimus, mis suudaks dünaamiliselt kohaneda teise vaimu (kunstliku või bioloogilise) tõenäoliste otsustega? Ettekanandes põhjendati, et see on võimalik ja et meil on juba palju teoreetilisi aluseid arendusprotsessi alustamiseks.

João Gluz jt (Brasiilia) vaatlesid tehisintellektis hästi tuntud arutlusmudelit BDI (ingl *belief-desire-intention*). Seda mudelit kasutatav tehiseagent peaks olema võimeline valima kõige ratsionaalsema tegevuse, mida saab teha aktsepteeritava aja jooksul, kui ressursid on piiratud ja teadmus pole täielik. Kuigi inimesed vajavad emotsioone, tegemaks otsuseid mittetäieliku info põhjal, ei võta traditsioonilised BDI mudelid arvesse agendi tundmuslikke seisundeid. Ettekanandes esitati BDI agentide jaoks emotsioonide hinna määramise rakendus, kasutades nn BDI keelt, mis ühendab loogikat ja tõenäosuslikku arutlust. Rakendus võimaldab teha vahet emotsioonide ja afektiivsete reaktsioonide vahel.

Mõned ettekanded olid pühendatud nn kognitiivsetele kaartidele (ingl *cognitive map*). Kognitiivne kaart on teadmuse esituse mudel graafina, kus sõlmed vastavad mõistetele ja servad esitavad nende mõistete vahelisi mõjusid. Igal mõjul on arvuline väärtus, mis seda iseloomustab. Aymeric Le Dorze jt (Prantsusmaa) ettekanandes esitati kaks kriteeriumi kognitiivse kaardi kvaliteedi hindamiseks: verifitseerimine, mis arvutab kaardi sisemise sidususe, ja testimine, mis väärtustab kaardi, arvestades kehtivaid kitsendusi. Need kriteeriumid näitavad, kas kaart sisaldab vastuolusid (ja on seega kõlbmatu) või mitte. Pakuti ka viis, kuidas kohendada neid kriteeriume võimalike väärtustega, mida mõju saab omada. Samade autorite teine ettekanne käsitles tõenäosuslikke kognitiivseid kaarte. Sellistes kaartides tõlgendatakse mõjude väärtusi tõenäosuslikena. Definieeriti formaalselt selle mudeli semantika. Pakuti välja operatsioon, kuidas

arvutada mõiste globaalset mõju mõnele teisele mõistele, nn tõenäolist levitatud mõju. Näitamaks, et mudel on kehtiv, pakuti protseduur kognitiivse tõenäosusliku kaardi esitamiseks Bayesi võrguna. Władysław Homenda jt (Poola) vaatlesid hägusaid kognitiivseid kaarte – abstraktseid nn pehmearvutuse (ingl *soft computing*) mudeleid, mida saab kasutada keeruliste süsteemide modelleerimiseks, kus esineb ebakindlus. Esitati kaks meetodit hägusa kognitiivse kaardi rekonstrueerimiseks, mis põhinevad masinõppel.

Dan Wu (Rootsi) käsitles ontoloogia-te integreerimist. Uurinud ontoloogiaid deponituumis, töötati välja kontekstireeglid. Iga selline reegel sisaldab tingimusi konteksti identifitseerimiseks. Need tingimused kirjeldatakse nn kontekstikriteeriumide abil, milleks võivad olla nt ontoloogia autor ja valdkond. Kui tingimused reeglis sobivad, siis reegel aktiveerub ja kontekstiinfo (reegli keha) sisestatakse arutlusmoodulisse, mida kasutatakse ontoloogia te integreerimiseks. Esineja tõi kontekstireegli konstrueerimise näite ja leidis, et kontekstiinfo kasutamise saavutatakse ontoloogia te integreerimisel paremaid tulemusi kui ilma selleta.

Oussama Ben Khiroun jt (Tuneesia) uurisid sõnatähenduste ühestamise mõju päringu laiendamisele ükskeelsel intelligentsel infootsingul. Ühestamiseks ja laiendamiseks seni pakutud lähenemised põhinevad korpuse analüüsil. Ettekandes esitatud mudel vastuse relevantsuse hindamiseks kasutab selle asemel nn võimaluste teooriat, mis mõõdab nii võimalikkust kui ka vajalikkust. Prantsuskeelse infootsingu hindamiseks tehti katsed standardse testikoguga ROMANSEVAL semantilisel ühestamisel ja CLEEF2003 tööpingiga päringu laiendamisel. Tulemused näitavad väidetavalt head mõju saagisele ja täpsusele.

M. Koit (Eesti) käsitles oma stendiettekandes eestikeelsete dialoogide automaattõtluse olemasolevaid vahendeid.

Mitmed ettekanded olid pühendatud suhtlusele arvutiga, st kuuluvad teise rühma. Michimasa Inaba jt (Jaapan) esitasid meetodi selliste dialoogiagentide (nt juturobotite) loomiseks, mis ei ole orienteeritud kindla ülesande lahendamisele. Agent valib olemasolevate lausungite hulgast sobiva, sõltuvalt kontekstist. Selleks omistatakse lausungitele sobivusjärgud, kasutades masinõpet. On välja töötatud ka õpiandmete kogumise keskkond. Eksperiment selle keskkonnaga on näidanud, et kontekstiinfo kasutamine parandab sobiva vastuse leidmist, kuigi olemasolevates süsteemides konteksti ei kasutata. Huma Shah jt (Suurbritannia) ettekanne käsitles Turingi testi, mis on inglise matemaatiku Alan Turingi 1950. aastal soovitatud meetod tehisintellekti olemasolu väljaselgitamiseks. Turingi testi saab täita kahel viisil: kolme osaleja test või kahe osaleja test. Esimesel juhul küsitleb inimesest kohtunik paralleelselt kahte tema eest varjatud katsealust ja otsustab, kumb neist on inimene ja kumb arvuti. Teisel juhul küsitleb kohtunik ühte varjatud katsealust ja otsustab, kas too on inimene või arvuti. Suhtlus on tekstipõhine ja toimub inglise keeles. Kui kohtunik ei suuda tuvastada arvutit, pidades seda inimeseks, siis tähendab see, et arvutil on (tehis)intellekt. Ettekandes tutvustati 2012. aastal Inglismaal läbi viidud testide tulemusi, kus rakendati mõlemat meetodit. Selgus, et kahe katsealuse paralleelne võrdlemine oli tehisintellekti tuvastamiseks kindlam test.

Jianfeng Xu jt (Jaapan) käsitlesid žestide ja kõne sünkroniseerimist multimodaalsetes dialoogiagentides. Sünkroniseerimise probleemi vaadeldi kui liikumise sünteesi probleemi. Esimesel sammul kasutati liikumisgraafi tehnikat, seades kitsendused žestide struktuurile, ja teisel sammul detailiseeriti seda struktuuri. Katsed on näidanud, et meetod annab parema tulemuse kui senised.

Mõned ettekanded käsitlesid kindlat tüüpi kasutaja modelleerimist, st kuu-

luvad kolmandasse rühma. Arman No-roozian jt (Holland) esitlesid autojuhtide treenimisel kasutatavat simuleerimis-raamistikku, mis võimaldab modelleerida individuaalseid autojuhte ja kiirtee liiklust, kus kasutatakse erinevaid juhtimisstiile. Yuki Katsura jt (Jaapan) tutvustasid argumenteerimispõhist tööriista, mis toetab olemasolevaid suhtlus-süsteeme, nagu Twitter jms, võimaldades põhjendatumat ja loogilisemat suhtlust inimeste vahel. Tööriist kasutab nn asünkroonset argumenteerimist, mis kasutab mitmevääruselise argumenteerimisloogikat. Seda võib nimetada ka asümptootiliseks ehk kasvavaks argumenteerimiseks, sest agent võib läheneda tõele samm-sammult, igal ajal hinnates talle esitatud argumente. Argumenteerimissüsteem on realiseeritud personaalsel tööriistal iPad.

Paljud ettekanded käsitlesid tehisintellekti teoreetilisi probleeme (tehisintellekti algoritmid, tehisnärviõrgud, andmekaeve, masinõppe meetodid jne) ja mitmesuguseid rakendusi, sh mobiilsed robotid, intelligentsed abivahendid erivajadustega kasutajatele, bussi- või rongiinfosüsteemid, meditsiinilise nõustamise süsteemid, virtuaalsed agendid veebis jpm.

Kokkuvõtteks võib tõdeda, et kõigil siin vaadeldud tehisintellektikonverentsidel pühendati loomuliku keele automaattöötlemisele, eriti keele mõistmise modelleerimisele märkimisväärselt palju tähelepanu. Olulised võtmemõisted, mis jäid kõlama kõigil konverentsidel ja viitavad aktuaalsetele uurimisprobleemidele, on suuremahuliste andmete

töötlemise ja andmekaeve, teadmuse, sh hoiakute ja tundmuste väljasõelumine (keele)andmetest. See pole ka üllatav, sest tehisintellektisüsteemid, mis on mõeldud tegutsema koos inimestega – see ei kehti kaugeltki kõigi tehisintellektisüsteemide kohta, on piisavalt neid, mis tegutsevad autonoomselt –, peavad inimestega ka suhtlema, ja seda inimese jaoks loomulikul viisil, nad peavad inimestest „aru saama”. Inimkeel esinebki kirjeldatud töödes mitte ainult ja mitte niivõrd andmete vahetamise vahendina, kuivõrd inimese mentaalsfääris toimuvate protsesside „lahtimuukimise” vahendina. Näiteks plahvatuslik areng valdkonnas, mida kokkuvõtvalt nimetatakse tundmusanalüüsiks, tõendab, kuivõrd inimpsüühika selle aspekti modelleerimist vaja on. Siin ei ole peamine keele formaalne struktuur, vaid semantilised ja pragmaatilised andmed, mis peituvad nii sõnade tähendusvarjundites kui ka lausete ja tekstide kogustruktuuris, suhtluskontekstis.

Kui tulla õeldu valguses eesti keele keeletehnoloogilise toe arendamise juurde, mida näeb ette nt eesti keele arengukava (2011–2017), siis on ilmne, et uues, järgnevas arengukavas peaks palju suurem rõhk olema vahetut keelekasutust (NB! mitte ainult suulist, sest nt säutsud ei ole suulised) kajastava materjali salvestamisel-töötlemisel. Seda enam, et see keelekasutuse aspekt ei ole universaalne, teiste keelte vastavatest andmekogudest vahetult ülevõetav, see on kultuurisidus.

HALDUR ÕIM, MARE KOIT