

INIMLOETAV JA MASINLOETAV: MÕISTETE ANALÜÜS

Priit Parmakson

Ülevaade

Infoühiskonnale on iseloomulik struktuuristatud, masintöötlemisele orienteeritud teabe osakaalu kasv, mille ilming on ka Eesti avalikus sektoris mitmes valdkonnas käsil olev selle-suunaline IT töö – XML-põhiste keeltele üleminek. Samas jääb info tarbijaks inimene. Teave peab muutuma nii masin- kui inimloetavaks, mis võib olla raskesti saavutatav ja tekitada pingeid infosüsteemide arendamisel ja kasutamisel. Esitatakse lihtne mudel, mille põhjal saab liigitada ja analüüsida inim- ja masinloetavate tekstide dünaamilises koosluses tekkivaid huvitavaid, kuid veel suhteliselt vähe uuritud keelenähtusi.

Võtmesõnad: inimloetavus, masinloetavus, struktuurne teave

Sissejuhatus

'Inimloetav' ja 'masinloetav' on kategooriate paar, millega on seotud laialdane ja huvitav, kuid samas suhteliselt vähe teoreetilist käsitlemist leidnud keelete nähtuste ring.

Igal soliidsel raamatul on ISBN kood. Tavaliselt on see trükitud tagakaanele ja koosneb triipkoodist ning triipkoodi dubleerivast 13-kohalisest numbrist. Triipkood on masinloetav, number on mõeldud inimesele lugemiseks (nt olukordades, kus triipkoodi skänner ei tööta). See lihtne näide on

omapärane selle poolest, et sama teave on esitatud rööbiti kahel kujul: üks orienteeritud arvutile, teine inimesele.

Digitaalsetes raamatutes on tõenäoliselt teisti. ISBN koodi triipkoodina esitada ei ole mõtet, sest teisele arvutile on koodi loomulik saata internetis. Mõistlik on hoida koodi raaltöödeldaval kujul ja kuvada seda inimesele vajadust mööda.

Masin- ja inimloetavusega seotud keeleliste nähtuste ring on päris lai. Neid ei ole eriti uuritud. Järgnevas vaadeldakse masin- ja inimloetavust nn struktuursele teabele (ingl *structured data*, sm *rakenteistettu tieto*) ülemineku kontekstis. Struktuurne teave on süstemaatilist infotöötlust, sealhulgas eriti arvutites töötlemist hõlbustavale kujule viidud teave. Teabe struktuuristamine¹⁰ hõlmab koodisüsteemide loomist ja teabe kodeerimist nende abil, vormingute kehtestamist, teabe märgendamist, metaandmete loomist jt tehnikaid. Tähtis eesmärk on vähendada ebavajalikuks peetavat keelelist varieeruvust sõnavara, süntaksi ja tähenduste standardimise abil. Tänapäeval tuntuim teabe struktuurse esitamise keel on XML¹¹. XMLi sarnaseid keeli on välja töötatud palju. XMLi tähtsaid omadusi on laiendatavus; see võimaldab luua üha uusi, üksteisega koosvõimelisi keelemoduleid (kasutatakse mõistet 'XML-põhine keel'). OASIS¹² ja W3C¹³ on kaks juhtivat struktuursete keelte loomisega tegelevat rahvusvahelist organisatsiooni.

¹⁰ Soomlastel on selle jaoks termin *tiedon rakenteistaminen*, vt nt: Lehmuskoski & Kuusisto-Niemi (2008).

¹¹ *Extensible Markup Language*, laiendatav märgendkeel. [NB! Käibel on ka terminid *märgenduskeel* ja *märgistuskeel*. Siingi oleks vaja terminikorrastuse käigus kokku leppida, milline kolmest on sobivaim. – Toim.]

¹² The Organization for the Advancement of Structured Information Standards, www.oasis-open.org.

¹³ World Wide Web Consortium, www.w3.org.

Teabe struktuuristatust käsitatakse sageli astmelisena: struktuuristatud (ingl *structured*), poolstruktuuristatud (ingl *semi-structured*), struktuuristamata (ingl *unstructured*).

Infoühiskonna praegune suund on struktuuristatud teabe osakaalu suurenemine. Loomuliku keele raaltöötlus nõuab tehisintellektitehnikaid, mis on kallid ja praegu veel ebarahuldavad. Praktilises infotöötuses on teabe struktuuristamine seevastu üldjuhul nii tõhus ja ökonoomne, et tänapäeva IT taseme juures peetakse seda vältimatuks: „Elektronilise asjaajamise vältimatu eeldus on, et dokumendid on struktuuristatud“ (Nieminen 2010: 32).

Üleminek struktuuristatud teabele

Eesti avalikus sektoris on käimas või juba lõppenud hulk ambitsioonikaid infotehnoloogilisi töid, mille tuumaks on üleminek struktuursele (struktuuristatud) teabele, eriti XML-põhiste keelte kaudu.

- Avaliku sektori asjaajamises on võetud strateegiliseks suunaks dokumentide muutmine XML keele põhisteks. Üleminek on järk-järguline, kuid pidev. Esialgu kirjeldatakse XML-põhiselt dokumentide metaandmed. Sellega muutuvad metaandmed hõlpsamini raaltöödeldavaks. Edaspidiseks nähakse ette XML-põhiste dokumentide üha laiemat kasutuselevõttu (Riigikantselei 2010).

- Õiguses minnakse üle XML-põhistele õigusaktidele (Majandus- ja Kommunikatsiooniministeerium 2007a: 15). Eesmärk on teisendada kõik Eesti seadused erilisse XML-keelde. Uusi seadusi hakatakse kirjutama spetsiaalse XML-tekstiredaktoriga, mis kontrollib sisestatava õigusteksti vastavust XML skeemiga kehtestatud reeglitele. Sellega muutuvad õigustekstid paremini raaltöödeldavaks, kuid inimene XML-keelset teksti

lugada ei suuda. Inimloetavaks teisendatakse XML-dokument nn stiililehe (transformatsiooni) rakendamisega. 2010. a käivitatus elektroonilise Riigi Teataja uus infosüsteem, kus kõik õigusaktid on viidud XML kujule (kasutajale esitatakse inimloetaval kujul).¹⁴

- Majandus- ja finantsaruandluses minnakse üle XBRL-keelele.¹⁵ See tähendab, et ettevõtte aruannet äriregistrile ei saa enam esitada vabas vormis. Formaalselt jääb tegevus küll eesti keelde, kuid eestikeelne teave hakatakse tükkideks ja pakendatakse masinloetavasse, rahvusvahelisse erilisse arvutikeelde. Toimunud on üllatav areng. Kümme aastat tagasi ei olnud raamatupidamise aruandele kehtestatud vorminõudeid.

- Riigi infosüsteemide semantilise koosvõime programm seab eesmärgiks teha avaliku sektori infosüsteemides käideldav teave kergemini seostatavaks ja taaskasutatavaks masinloetavate ontoloogiate abil XML-põhises OWL keeles (Majandus- ja Kommunikatsiooniministeerium 2007b).

Mitmes nendest töödest on Eestis võrdluses teiste riikidega saavutatud väljapaistvaid tulemusi. Väga vähe on riike, kus kõik õigusaktid on struktuurse teabe (XML) kujul. Kuid samas suunas kulgeb areng ka teistes riikides.

Need tööd muudavad mitmes valdkonnas keelelist situatsiooni, eelkõige masinloetavate ja inimloetavate keelte ning keelevormide vahelkorda.

XML ja teised struktuurse teabe keeled on suunatud arvutite vahelisele infoedastusele. XML-põhine tekst ei ole mõeldud otse inimesele esitamiseks. Seetõttu on õigustatud nende keelte nimetamine masinloetavateks keelteks (lühendatud

¹⁴ Vt: <https://www.riigiteataja.ee/abiLeht.html?id=1>.

¹⁵ XBRL töö lehekülj: <https://ajaveeb.just.ee/e-aruandlus/>; taksonoomiatega saab tutvuda: <http://xbml.eesti.ee>.

masinkeelteks). Kuid siiski puutub inimene asjasse ja seda kahel moel: inimesele infosüsteemides esitatav teave pärineb XML-põhistest tekstidest (rakendatakse inimloetavust parandavaid teisendusi ja kujundusvorminguid); teiseks, XML-põhiste tekstidega töötavad infosüsteemide arendajad ja liidestajad.

Just selles kontekstis tekib infosüsteeme arendades, aga samuti kasutades, edusammude kõrval ka pingeid, mis on põhjustatud sellest, et inim- ja masinloetavuse eesmärkide võrdsel määral saavutamine on tihti problemaatiline. Üks ja seesama teave peab sageli olema kättesaadav nii masinloetaval kui ka inimloetaval kujul. Näiteks Eesti avalikus sektoris kasutatavad klassifikaatorid peavad olema kättesaadavad XML-formaadis.¹⁶ Praegu ei ole kõik klassifikaatorid veel XML-vormingus. Kasutatakse Exceli tabeli jms masintöötuseks liiga vähe struktureeritud esitust, sest neid on inimesel lihtsam teha ja lugeda. Samas peab ka raaltöötuseks sobivat, XML-vormingus klassifikaatorit sageli ikkagi inimkasutajale sobivas, inimloetavas vormingus esitama.

Loetavuse mõiste modelleerimine

Infosüsteemides (ja infoühiskonnas laiemalt), kus teabe tootjaks ja kasutajaks on nii inimesed kui arvutid, tekib sageli vajadus esitada sama teave üheaegselt kahel kujul. Siit tuleneb vajadus analüüsida ja võimalusel modelleerida inimloetavuse ja masinloetavuse mõistet. Ilmne ja triviaalne on teabe loetavuse suhtelisus, ajast ja kontekstist sõltuvus. Inimloetavus sõltub lugeja eelteadmiste, kognitiivsete võimete ja

¹⁶ Klassifikaatorite süsteemi kehtestamine. Vabariigi Valitsuse 09.10.2003. a määrus nr 255, § 4 lg 3, <https://www.riigiteataja.ee/akt/632688>.

lugemisoskuse kõrval suuresti ka kultuurist. Masinloetavus sõltub tehnoloogia arengust. Tehisintellekttehnikaid sisaldav tekstituvastustarkvara võimaldab lugeda materjali, mida madalama taseme tehnoloogia korral peetaks töödeldamatuks. Praegune tehnika tase ei võimalda automaatselt töödelda PDF-failina saadetud arvet; automatiseeritud töötamiseks peab arve olema e-arve standardi kohases vormingus.

Tegu on niisiis mõistetega, mille täpne defineeriminegi on raske.¹⁷ Tegelikuses tekib tihti vajadus objektiivselt hinnata, kas ja kuidas tekst on loetav. Näiteks kas programm on piisavalt dokumenteeritud, s.t kas programm inimesele (teisele programmeerijale) on loetav. Arvutite-eelsel ajal, peamiselt pedagoogilise kirjanduse valiku eesmärkidel, on välja töötatud mitmesuguseid teksti loetavuse mõõdikuid¹⁸, kuid digitaalses keskkonnas eeldab nende kasutamine vähemalt suurt kohandamist (Sallis & Kassabova 2000)¹⁹. Väga oluline uus uuringute ja infosüsteemide loomise suund on (veebi) kasutatavus (ingl *usability*); sellega on seotud veebi käideldavuse teema.²⁰

¹⁷ Termin *masinloetav* esineb 13 Eesti õigusaktis (<https://www.riigiteataja.ee/>), peamiselt isikut tõendavatele dokumentidele kantava teabe kontekstis. Uueks kasutuseks on raamatupidamise taksonoomia masinloetava vormi määramine. Inimloetavuse mõistet ilmutatud kujul Eesti õiguses ei kasutata, küll aga Euroopa Liidu õiguses: „kasutajaliideses kasutatav kihi inimkeelne loetav pealkiri“, „inimloetav tähis või nimi“ (eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:01:ET:HTML).

¹⁸ Vt nt: (Bailin & Grafstein 2001) , aga samuti “Readability” artikkel *Wikipedia*. [Eestis nt Jaan Mikk – Toim.]

¹⁹ Siiski, Google'i otsimootor võimaldab 2010. a alates otsingu tulemusi väljastada loetavustasemete (*readability level*) järgi liigitatult. Vt: <http://www.seroundtable.com/google-reading-level-algorithm-2638.html>.

²⁰ Web Accessibility Initiative (WAI). www.w3.org/WAI/.

Inimesele tähendab loetavus kindlasti mõistetavust. Arusaamise sügavus on mõiste oluline dimensioon. Pinnapealsel tasandil lugemise võimalusest ju ei piisa, vajalik on piisavalt sügav arusaamine. Samas tuleb loetavust mõista ka kui kasutatavust. Oluline ei ole ainult tulemus (arusaamine), vaid ka ajakulu ja kognitiivne jõupingutus, mis inimesel teksti mõistmiseks tuleb teha. Eichelberger ja Schmid (2009) on loonud UML diagrammide (infotehnoloogiline modelleerimiskeel) "esteetilise kvaliteedi" nõuete mudeli. Esteetilise kvaliteedi all mõistavad nad eelkõige loetavust.

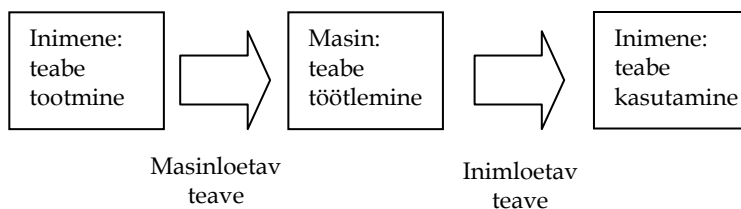
Masina puhul tähendab loetavus seda, et masin saab teavet mingil kasulikul viisil töödelda. Raaltöödeldavus tähendab tekstis piisava struktuuri – ja tähtis on, et just raaltöötluks tarviliku struktuuri – olemasolu.

Loetavuse analüüsil ei ole õigustatud piirduda ainult teksti kasutamisega, vaid vaadelda tuleks teabe elutsükli tervikuna (joonis 1). Teabe tootmine ja teabe kasutamine on kaks peamist järku; mõlemaga võivad tänapäeval tegeleda inimesed ja arvutid. Tihti on vaja teisendada teave inimloetavast vormist masinloetavasse või vastupidi. Teabe tootmise, teisendamise ja kasutamise vahendid ning tingimused moodustavad infrastruktuuri, mis määrab loetavuse nõuete rahuldamise.

Teabe masinloetaval kujul tootmine või olemasoleva teabe masinloetavaks teisendamine, kui seda peavad tegema inimesed, võib struktuursele teabele ülemineku töös kujuneda pudelikaelaks. Soome kohtute elektroonilisele asjaajamisele üleminekut käsitlevas uurimuses tõdetakse: „Kui dokumendihalduses võetakse arvesse struktuuristamine, metaandmed ja pääsuõigused, siis muutub dokumentide tootmine vältimatult vähem kasutajasõbralikuks. Dokumentide struktuuristamise

takistuseks on varem olnud see, et kohtud ei ole soostunud kasutama struktuuristamist võimaldavaid abivahendeid, vaid on tahtnud dokumente toota vabas vormis.” (Nieminen 2010: 32).

Eestis on põrnutud samasugustele raskustele. Eesti esimeses kohtulahendite infosüsteemis KOLA kujunes probleemiks andmete kandmine süsteemi, sest kohtusekretärid pidid lahendite tekstidest süsteemile vajalikke metaandmeid käsitsi välja otsima (Laks 2003).



Teabe elutsükkel: tootmine → töötlemine → kasutamine.

Kõigis kolmes positsioonis võib olla nii inimene kui arvuti. Liikuv teave võib olla inimloetav, masinloetav või mõlemat.

Joonis 1. Teabe elutsükkel.

Vaja on nii inimloetavust (arusaadav inimestele) kui masinloetavust (raaltöötlus peab olema võimalik). Kaks peamist teed selle saavutamiseks on: 1) luua kõrvuti kaks esitust – inimloetav ja masinloetav; 2) leida esitus, mis oleks arusaadav nii inimesele kui arvutile (kui see on võimalik). Samuti on tähelepanuväärne idee lasta raalil genereerida inimloetav kujud masinloetava alusel.

Erinev esitus masinale ja inimesele

See lihtne idee on internetitehnoloogiates laialt kasutusel: “Machines should get RDF data and humans should get a readable representation, such as HTML.” (W3C 2008).

Ontoloogiate kirjeldamise keelel OWL (versioon 2) on viis süntaktilist kuju, neist kaks on suunatud inimesele, kolm raaltöötlusele (W3C 2009). Struktuurse teabe kirjeldamise keel RELAX NG pakub nn kompaktset süntaksivarianti, mille eesmärk on maksimeerida inimloetavust (OASIS 2002). Info- tehnoloogilise modelleerimiskeele UML „inimnäoks“ on visuaalne diagrammikeel, masinale on aga suunatud XML-põhine süntaks (XMI keel). Mistahes rööpesituse nõrk koht on ebamajanduslikkus (mitmesse keelde paneku kulukus) ja erinevate esituste sünkroonimise vajadus. Inimesed ja arvutid ei tarvitse ka sisemiselt moodustada keekekasutuse seisukohalt homogeenseid gruppe – tekib tarvidus enamate keelte järele.

Digitaalses keskkonnas ringlevaid tekste iseloomustab sageli dünaamilisus: tekstist moodustatakse aja jooksul uusi versioone. Siis kerkib inim- ja masinloetavate kujude sünkroonimise probleem. Sama teabe inim- ja masinloetava esituse käsitsi sünkroonimine võib äärmise töömahukuse tõttu osutada võimatuks. Seetõttu otsitakse abi tarkvarast. See võib töötada ühes või teises suunas (dokumentatsiooni genereerimine programmikoodi põhjal; masinloetavas vormingus teabe generaatorid) või olla mõlemapidised. Näiteks Braband jt (2008) leiavad, et XML on inimese jaoks liiga sõnaohter (ingl *verbose*) ja pakuvad välja kahe teineteisega sünkroonitud süntaksi (inimloetav ja tavaline XML) süsteemi. Sünkroonimine on kõige lihtsam, kui suudetakse programmeerida automatiseeritud teisendus ühest esituskeelest teise. Seda

strateegiat kasutatakse veebis: veebilehe sisemiseks esituseks on masinloetav XML. Inimloetav, veebilehitsejas kuvatav esitus saadakse transformatsioone ja laadilehti (CSS keeles kirjeldatud teisendusreeglid) rakendades.

Teabe elutsükklisse kuulub ka teabe säilitamine. Oluline tendents on teabe säilitamine ainult masinloetaval kujul. Näiteks elektroonilises Riigi Teatajas säilitatakse õigusakte masinloetavas XML-vormingus. Igale konkreetsele kasutajale genereeritakse inimloetav õigusakti esitus (selleks kasutatakse XSLT ja CSS tehnoloogiat). Sellesse skeemi on sisse ehitatud eeldus, et inimloetav on alati genereeritav masinloetavast.²¹

Tekstid, mis on ühtlasi inim- ja masinloetavad

Ideaalne oleks, kui tekst oleks mõistetav (töödeldav) nii inimesele kui arvutile.²² Selliste nõuetega tekstide suure klassi moodustavad programmid. Programm peab olema nii masintöödeldav (arvutiga täidetav) kui ka inimese jaoks loetav (sest programme modifitseerivad (veel) inimesed). Programmi üheaegse masin- ja inimloetavuse saavutamiseks on mitu teed. Levinuim moodus on inimloetava teksti (kommentaaride) lisamine masintöödeldava teksti (programmikoodi) vahele. Sümmeetria printsüüpi rakendades võime käsitleda seda ka raaltöödeldava teksti lisamisena inimloetava

²¹ Võib küsida: kumb on primaarne: inim- või masinloetav vorm?

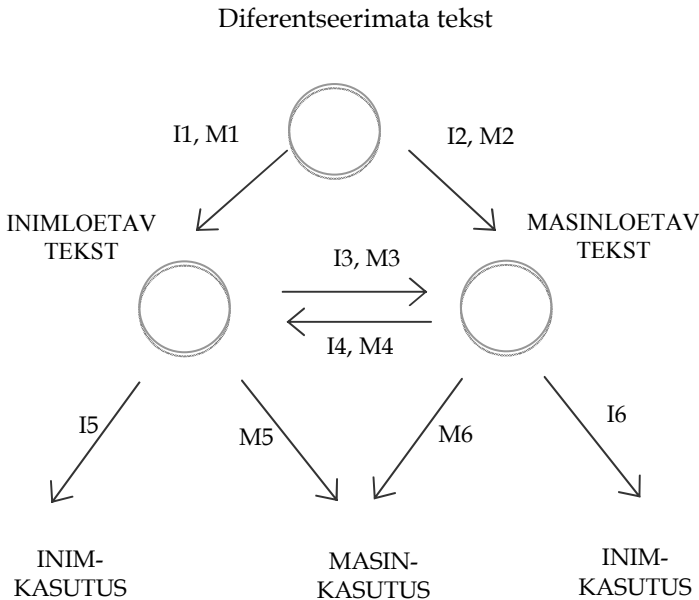
²² Vahel on see paratamatult vajalik. Näiteks veebilehtede aadresse (tehnilises kõnepruugis URIsid (*Uniform Resource Identifier*)) kasutavad nii inimesed kui raalid. Inimloetavat URId on aga tihti raske saavutada, sest inimloetavus võib minna vastuollu raalisüsteemides realiseeritud töötlus-põhimõtetega.

teksti sisse. Kommenteeritud programmid on sisuliselt kakskeelsed tekstid, kusjuures keeltepaaris ei ole kaks inimkeelt, vaid inim- ja masinkeel. Tehtud on otseselt inimloetavate programmeerimiskeelte loomise katseid. Lähenemised on olnud rohkem või vähem radikaalsed, alates inimkeelepärase sõnavara valikust inimkeele lauseehitust jälgendavate konstruktsioonideni ja kuni inimkeele piiratud alamhulkade lubamiseni. Radikaalseim seisukoht on püüda raal loomulikku keelt töötlemata panna. Selles suunas on edu saavutatud päringukeeltes: päringu võib anda loomulikus keeles; seejuures ei ole siiski tagatud, et raal päringut õigesti ja täielikult mõistab. Isedokumenteeriv kood on programmeerijate keskkonnas mõnel määral levinud radikaalne vaade – uskumus, et valides klasside, muutujate jm objektide nimed väljendusrikkad, on kood suure osas iseennast seletav (inimesele). Seega nii masin- kui ka inimloetav.

Raaltöödeldavate tekstide inimloetavus puudutab suhteliselt kitsast ringkonda – infosüsteemide arendajaid, kes programme, mudeleid ja arvutiga töödeldavaid andmeid kirjutama, lugema ja modifitseerima peavad. Huvitav küsimus on XML keele inimloetavus. XML on inimloetav(am) võrreldes arvutites andmete esitamiseks kasutatava nn binaarkoodiga (nullid ja ühed). XML on märgendkeel; võib väita, et XML-vormingus tekst on isegi loetavam kui tavaline tekst, sest teksti on pikitud teksti struktuuri kirjeldavaid märgendeid.

Inim- ja masinloetavuse mudel

Eelkäsitletust ilmneb, et inim- ja masinloetavuse mõisteid juba kasutatakse aktiivselt, ent nende laiem kasutamine võimaldaks nähtusi veel paremini avada ja rakendada. Kas neid nähtusi saaks süsteemsemalt ja kompaksemalt esitada? Pakun välja järgmise, käsitletud nähtusi süstematiseeriva ja üldistava mudeli. Mudelis on nelja tüüpi elemente: tekstid (diferentseerimata tekst, inimloetav tekst, masinloetav tekst), aktorid (I - inimene, M - masin), teksti paremat loetavust edendavad tegevused (I1, M1, I2, M2, I3, M3, I4, M4) ja teksti kasutamine (I5, I6, M5, M6).



Joonis 2. Inim- ja masinloetavuse mudel.

Elementidevahelised tähtsamad seosed on esitatavad järgmiste lausetega.

1) Inimloetavus ja masinloetavus on muutunud tekstide oluliseks omaduseks. Sageli on vaja välja selgitada, kas (diferentseerimata) tekst on inim- ja/või masinloetav.

2) Inimloetavuse, aga ka masinloetavuse nõutava taseme saavutamiseks on sageli vaja teksti spetsiaalselt töödelda või teisendada (diferentseerimata tekst → inimloetav tekst; diferentseerimata tekst → masinloetav tekst).

3) Üks ja sama tekst võib põhimõtteliselt olla nii masin- kui ka inimloetav (nt isedokumenteeruv programmikood). Elus jääb see siiski sageli saavutamatuks ideaaliks. Seetõttu leiavad kasutust sama teabe erinevad, inim- ja masinkasutajale suunatud versioonid. Näiteks e-kirjale võib lisada vCard vormingus elektroonilise visiitkaardi²³, kuid ilma spetsiaalse programmita ei saa inimene seda kaarti tavapärasel vormingus lugeda.

4) Tegelikult esineb juhtumeid, kus inimloetavale tekstile on vaja anda masinloetav kuju. Vastava teisenduse võib teha inimene (I3), teatud juhtudel on võimalik teisendust osaliselt või täielikult automatiseerida (M3).

5) Harvad pole ka juhud, kus masinloetavast tekstist on vaja moodustada inimloetav tekst (I4, M4).

6) Kui ühe ja sama teabe inim- ja masinloetavat versiooni hoida ja kasutada, võivad tekkida sünkronimise ja ajakohastamise, aga samuti semantilise vastavuse probleemid.

²³ vCard. <http://www.imc.org/pdi/>.

7) Inimloetavat teksti saab inimene normaalselt lugeda (I5). Masinloetav tekst on mõeldud raaltöötluks (M6). Nende normaaljuhtude kõrval pole harvad ka olukorrad, kus inimloetavat (kuid mitte masinloetavat) teksti on vaja masinaga kasutada (M5) ja vastupidi – inimene on sunnitud kasutama algselt masinaga töötlemiseks mõeldud teksti (I6).

Mudel võimaldab jälgida mõningaid inim- ja masinloetavuse mõistepaariga seotud tähtsamaid kommunikatsioonielemente ja -strateegiaid. Kuid sellega käsitusala nähtuste keerukus ei ammendu. Teabe ettevalmistamise toimingud ja teabeesituse abistavad aspektid (nagu näiteks tehniline vorming) võivad kommunikatsiooniprotsessides omandada esialgselt sisust lahknevaid kõrvaltähendusi (Feldman & March 1981). Masinloetavas vormingus teksti saatmine inimesele võib kommunikatsiooniaktina kanda teate sisust erinevat tähendust (nt demonstreerida saatja IT alast pädevust).

Kokkuvõte

Infoühiskonnas on teabe tootjateks, töötlejateks ja kasutajateks nii inimesed kui arvutid. Teabe raaltöödeldavuse saavutamine ei tohiks toimuda inimloetavuse arvelt. Seda on praeguste tehnoloogiatega raske saavutada ja vahel tuleb teha kompromisse. Sellelt pinnalt tekkivaid keelelisi konflikte on mõtet tähelepanelikult uurida ja lahendada.

Kirjandus

- Bailin, Ann & Grafstein, Alan 2001. The linguistic assumptions underlying readability formulae: a critique. – *Language & Communication*, vol 21, 285–301.
- Brabrand, Claus & Møller, Anders & Schwartzbach, Michael 2008. Dual syntax for XML languages. – *Information Systems*, vol 33, no 4–5, 385–406.
- Eichelberger, Holger & Schmid, Klaus 2009. Guidelines on the aesthetic quality of UML class diagrams. – *Information and Software Technology*, vol 51, 1686–1698.
- Feldman, Martha & March, James 1981. Information in Organizations as Signal and Symbol. – *Administrative Science Quarterly*, vol 26, no 2, 171–186.
- Laks, Mihkel 2003. Andmekvaliteedi hindamine kohtulahendite registri näitel. Magistritöö. Tallinna Ülikool. http://www.cs.tlu.ee/osakond/opilaste_tood/magistri_tood/2004_kevad/Mihkel_Laks/Mihkel_Laks_Magistri_Too_Lisadeta.pdf.
- Lehmuskoski, Antero & Kuusisto-Niemi, Sirpa 2008. Sosiaaliaan sanasto asiakastietojärjestelmää varten. Sosiaali- ja terveystieteiden tutkimus- ja kehittämiskeskus (Stakes). www.stakes.fi/verkojulkaisut/raportit/R30-2008-VERKKO.pdf.
- Majandus- ja Kommunikatsiooniministeerium 2007a. XML-põhise õigusaktide ettevalmistamise ja publitseerimise keskkonna loomine. Infoühiskonna arengukava 2013. www.valitsus.ee/failid/infyhisk_rak_plaan_uus.pdf.
- Majandus- ja Kommunikatsiooniministeerium 2007b. Semantilise koosvõime raamistik. <http://www.riso.ee/et/et/koosvoime/semantika/raamistik>.
- Nieminen, Pertti jt 2010. Yleisten tuomioistuinten asianhallinnan kehittäminen. Oikeusministeriö. ISBN 978-952-466-741-8. www.om.fi.
- OASIS [Organization for the Advancement of Structured Information Standards] 2002. RELAX NG Compact Syntax. www.relaxng.org/compact.html.

- Riigikantselei 2010. Juhis dokumendiliigi XML andmekirjelduse koostamiseks. <http://valitsus.ee/et/riigikantselei/dokumendihaldus/regulatsioon/juhised>.
- Sallis, Philip & Kassabova, Diana 2000. Computer-mediated communication: experiments with e-mail readability. - Information Sciences, vol 123, 43-53.
- W3C 2008. Cool URIs for The Semantic Web. www.w3.org/TR/cooluris/.
- W3C 2009. OWL 2 Web Ontology Language. Document Overview. www.w3.org/TR/owl2-overview/.

Human-readable and machine-readable: conceptual analysis

Priit Parmakson

Summary

Increase of structured data, motivated by efficiency of machine processing (computing), is a characteristic trend of information society. In Estonian public sector too, a number of IT projects are under way in this direction (transfer to XML languages). However, humans still remain significant if not major users of information. Information has to become both machine-readable and human-readable. This is often hard to achieve and may put a lot of stress on development and use of information systems. A simple model is presented that can help categorise and analyse linguistic phenomena that arise in dynamic co-existence of human and machine readable texts.

Keywords: human readability, machine readability, structured data.

Autor

MA Priit Parmakson on Riigi Infosüsteemi Ameti riigi infosüsteemi haldussüsteemi osakonna analüütik.
priit.parmakson@ria.ee