

Að minnsta kosti 21 Evrópumál á stafrænan dauða á hættu

Rannsókn gerð af fremstu máltækni sérfræðingum Evrópu bendir til að flest Evrópumál, þar á meðal íslenska, séu í útrýmingarhættu á stafrænni öld

Flest Evrópumál, þar á meðal íslenska, eiga á hættu stafrænan dauða samkvæmt nýrri rannsókn fremstu sérfræðinga Evrópu á sviði máltækni. Sérfræðingarnir lögðu mat á stöðu máltækni fyrir 30 af um 80 tungumálum álfunnar og komust að þeirri niðurstöðu að stafrænn stuðningur við 21 af þessum 30 málum væri lítill sem enginn. Rannsóknin var gerð á vegum META-NET, evrópsks öndvegisnets sem tekur til 60 rannsóknarsetra í 34 löndum, þ. á m. Íslandi.

Rannsóknin var unnin af meira en 200 sérfræðingum og niðurstöðurnar birtar í Hvítbókaröð META-NET, 30 binda ritsafni sem er bæði prentað og á netinu (<http://www.meta-net.eu/whitepapers>). Mat var lagt á máltækni stuðning við sérhvert tungumál á fjórum mismunandi sviðum, sem eru vélþýðingar, talsamskipti, textagreining og aðgengi að mállegum gagnasöfnum. Að mati sérfræðinganna fellur 21 af þessum 30 tungumálum (70%) í neðsta flokk, þar sem er „lítill sem enginn stuðningur“, á a.m.k. einu af þessum sviðum. Nokkur tungumál, m.a. íslenska, en einnig lettneska, litháíska og maltneska, fá þessa einkunn á öllum sviðunum fjórum. Íslenska lenti reyndar í næstneðsta sæti af málunum 30 – aðeins maltneska stendur verr að vígi. Ekkert tungumálanna var talið njóta „afbúðagóðs stuðnings“, og aðeins enska taldist hafa „góðan stuðning“. Þar á eftir komu tungumál eins og franska, hollenska, ítalska, spænska og þýska með „sæmilegan stuðning“. Tungumál eins og baskneska, búlgarska, gríska, katalónska, pólska og ungverska njóta „brotakennds stuðnings“, og lenda því í áhættuhópi.

„Niðurstöður rannsókna okkar eru verulegt áhyggjuefni. Meirihluti Evrópumála býr við alvarlegan skort á málföngum – mállegum gagnasöfnum og máltækni búnaði – og sum eru nánast vanrækt tæknilega. Því má segja að mörg af tungumálum okkar séu ekki enn búin undir framtíðina“, segir Hans Uszkoreit prófessor, stjórnandi META-NET, vísindalegur stjórnandi hjá DFKI (Þýsku rannsóknarmiðstöðinni í gervigreind, Deutsches Forschungszentrum für Künstliche Intelligenz) og annar ritstjóri hvítbókaðarinnar. Hinn ritstjórinn, dr. Georg Rehm hjá DFKI, bætir við: „Það er gífurlegur munur á máltækni stuðningi milli evrópskra tungumála og tækni sviða. Bilið á milli ‘stóru’ og ‘litlu’ tungumálanna breiðkar sífellt. Við verðum að sjá öllum litlum og málfangafátækum tungumálum fyrir nauðsynlegri grunntækni – að öðrum kosti bíður þeirra stafrænn dauði.“

Innan máltækni er smíðaður hugbúnaður sem getur unnið með mannlegt mál, bæði talað og ritað. Meðal velþekktra dæma um máltækni búnað eru villuleitarforrit fyrir stafsetningu og málfar, gagnvirk þjónusta í snjallsímum (eins og Siri í iPhone), samræðukerfi sem vinna í gegnum síma, vélrænar þýðingar, vefleitarvélar og talgervlar í leiðsögukerfi bíla. Nútíma máltækni búnaður byggist einkum á tölfærðilegum aðferðum sem þjálfar þarf á gífurlegu magni af rituðu eða töluðu máli. Það er sérlega erfitt að afla nægilega mikilla gagna fyrir tungumál sem fáir tala. Máltækni búnaður sem byggist á tölfærðilegum aðferðum er líka háður ýmsum takmörkunum eins og t.d. má sjá á rangþýðingum vélþýðingarkerfa á netinu, sem oft geta þó verið brotlegar.

Evrópuríkjum hefur tekist að ryðja úr vegi flestum hindrunum fyrir samstarfi milli þjóða álfunnar. Ein hindrun er þó enn til staðar og virðist óyfirtíðanleg: Ósýnilegir tungumálaþröskuldar sem torvelda frjálst flæði þekkingar og upplýsinga. Þeir standa einnig í vegi fyrir því langtíamarkmiði að koma upp sameiginlegum stafrænum markaði vegna þess að þeir hindra frjálst flæði varnings, framleiðsluvara og þjónustu. Þótt máltækni hafi burði til að ryðja tungumálaþröskuldum úr vegi með nútíma vélþýðingarkerfum sýna niðurstöður úr rannsókn META-NET að mörg Evrópumál eru ekki enn undir þetta búin. Það eru alvarlegar eyður í tækninni vegna þess hve rannsóknar- og þróunarstarf er oft enskumiðað, vegna skorts á stuðningi og fjármagni, og einnig vegna þess að skýra sýn í rannsóknum og tækniþróun skortir.

Nauðsynlegt er að hefja viðamikil samræmt átak í Evrópu til að byggja upp þann tækni búnað sem vantar, og yfirfæra tæknina á meirihluta Evrópumála. Sterk rök eru fyrir því að ráðast til atlögu við þetta risavaxna verkefni í sameiginlegu átaki ríkja Evrópusambandsins og Evrópska efnahagssvæðisins, með þátttöku iðnaðarins. Þessi rök felast m.a. í háum kostnaði smærri málsvæða á hvern íbúa; nauðsyn á yfirfærslu tækni milli tungumála; skorti á samvinnu og samnýtingu málfanga, búnaðar og þjónustu; og þeirri staðreynd að mörk málsvæða fara ekki alltaf saman við landamæri ríkja. Evrópuríki verða að bregðast við og búa tungumál sín undir stafræna öld. Þau eru ómetanlegur þáttur menningararfs okkar og eiga því skilið að framtíð þeirra sé tryggð. Hinn 26. september er Evrópski tungumáladagurinn haldinn að frumkvæði Evrópuráðsins til að minna á mikilvægi þess að hlúa að og þróa áfram hina auðugu mállegu og menningarlegu arfleifð álfunnar okkar. Niðurstöður META-NET eru holl áminning um þær ögranir og þau tækifæri sem mállegur arfur okkar stendur frammi fyrir á upplýsingaröld.

Máltækni: Bakgrunnur

Máltækni nýtist nú þegar við ýmis hversdagsleg verk, svo sem að skrifa tölvupóst eða kaupa miða á netinu. Víða um heim hagnýtir fólk máltækni þegar það leitar að vefsíðum og þýðir þær, notar villuleitarforrit í ritvinnslukerfum, notar raddskipanir í símanum eða til að stjórna afþreyingarþúnaði í bílnum, fær tillögur frá netbókabúðinni sinni, eða fylgir töluðum leiðbeiningum leiðsagnarkerfis í síma. En fæst af þessu er til fyrir íslensku, og verður það ekki nema gripið sé til aðgerða. Í náginni framtíð munu menn geta talað við ýmis tölvuforrit, vélar og tæki, þar á meðal langþráð þjónustuvélmenni sem brátt munu sinna margvíslegum störfum á heimilum og vinnustöðum. Þegar fólk þarf á leiðbeiningum að halda mun það einfaldlega biðja um þær, hvar sem það er statt, og þegar fólk þarfnast aðstoðar mun það óska eftir henni – upphátt. Veröldin breytist þegar samskiptahindranir milli mannfólksins og tækninnar hverfa.

Það er almennt viðurkennt að máltækni sé einn helsti vaxtarbroddur upplýsingatækninnar um þessar mundir. Alþjóðleg stórfyrirtæki eins og Google, Microsoft, IBM og Nuance hafa ráðist í miklar fjárfestingar á þessu sviði. Í Evrópu hafa hundruð smárra og meðalstórra fyrirtækja sérhæft sig í ýmsum máltækniþúnaði eða þjónustu. Máltækni gerir fólki kleift að vinna saman, læra, stunda viðskipti og skiptast á þekkingu yfir mörk tungumála og óháð tölvukunnáttu.

Hvítbókaröð META-NET

Hvítbókaröð META-NET, „Tungumál í evrópsku upplýsingasamfélagi“ („Languages in the European Information Society“), lýsir stöðu máltækni í 30 Evrópumálum og gerir grein fyrir bráðustu hættum og brýnustu verkefnum hvers málsvæðis. Ritroðin tekur til allra opinberra tungumála Evrópusambandsins og Evrópska efnahagssvæðisins og nokkurra annarra mála sem töluð eru í álfunni. Þótt áður hafi verið gerðar mikilvægar og víðtækar rannsóknir á tilteknum þáttum tungumála og tækni hefur fram til þessa ekki verið til neitt aðgengilegt safnrit sem kynnir meginniðurstöður og ógranir sem hvert tungumál stendur frammi fyrir gagnvart tæknivæddri margmála Evrópu. Hvítbókaröð META-NET fyllir þetta skarð. META-NET getur nú sýnt fram á hvers vegna flest tungumálin standa frammi fyrir alvarlegum vandamálum og bent á uggvænlegustu eyðurnar. Samtals tóku meira en 200 höfundar þátt í samningu hvítbókanna eða lögðu til efni.

Eftirtalin Evrópumál eru tekin fyrir í hvítbókaröðinni: Baskneska, búlgarska, danska, eistneska, enska, finnska, franska, galísíska, gríska, hollenska, írski, íslenska, ítalska, katalónska, króatíska, lettneska, litháíska, maltneska, norska (bókmál og nýnorska), portúgalska, pólska, rúmenska, serbneska, slóvakíska, slóvenska, spænska, sænska, tékkneska, ungerska og þýska. Hver hvítbók er rituð á því máli sem hún fjallar um, og hefur einnig að geyma heildarþýðingu textans á ensku.

Um META-NET og META

META-NET er öndvegisnet 60 rannsóknarsetra í 34 löndum, þ. á m. Íslandi, helgað uppbyggingu tæknilegra forsendna fyrir margmála upplýsingasamfélagi í Evrópu. META-NET er að hluta fjármagnað af Evrópusambandinu gegnum fjögur verkefni, þ. á m. META-NORD sem Ísland tekur þátt í.

META-NET stendur á bak við META, margmála tæknibandalag Evrópu (Multilingual Europe Technology Alliance). Meira en 600 aðilar frá 55 löndum, þ. á m. rannsóknarsetur, háskólar, smá og meðalstór fyrirtæki, og allmörg stórfyrirtæki hafa nú þegar gengið til liðs við þetta opna tæknibandalag.