

Reykjavík, 19. mars 2018

Til allsherjar- og menntamálanefndar Alþingis

Efni: Umsögn um 36. mál, Frumvarp til laga um breytingu á höfundalögum

Á seinustu árum hefur orðið gerbreyting á aðferðum við gerð máltæknibúnaðar þannig að í stað málfræðireglna og mynstra er nú mest byggt á gervigreind og tauganetum (e. neural networks). Í því felst að öflugar tölvur eru látnar lesa mikið magn mállegra gagna – ritaða texta eða upptök-ur af töluðu máli – og læra af þeim. Tölvurnar finna þá sjálfar mynstur í gögnunum, í stað þess að mynstrin eða reglurnar séu skrifuð fyrir þær.

Þessar aðferðir hafa ýmsa kosti umfram eldri aðferðir og skila betri árangri. Fyrir lítið málsvæði eins og það íslenska skiptir þó kannski mestu máli að aðferðirnar eru óháðar tungumálum þar sem ekki þarf að skrifa sérstakar reglur eða mynstur fyrir hvert tungumál. Tölvurnar læra jafnauðveldlega af gögnunum á hvaða tungumáli sem þau eru.

Til að unnt sé að beita þessum aðferðum þarf þó ein grundvallarforsenda að vera fyrir hendi: Málleg gögn, og mikið af þeim. Því meiri sem gögnin eru, þeim mun meiri og nákvæmari upplýsingar geta tölvurnar dregið úr þeim, og þeim mun betri máltæknibúnað er hægt að þróa á grundvelli þessara upplýsinga. Þetta á við um allar tegundir máltæknibúnaðar – talgreina, talgervla, þýðingarforrit, leiðréttingarforrit o.s.frv.

Málleg gögn fyrir íslensku eru hins vegar af skornum skammti, miðað við tungumál fjölmennari þjóða. Þess vegna skiptir gífurlegu máli fyrir íslenska máltækni og þróun íslensks máltæknibúnaðar að hægt verði að nýta öll tiltæk gögn við slíka upplýsingavinnslu. Það frumvarp sem hér liggur fyrir er því mjög mikilsvert fyrir íslenska máltækni og stöðu íslensku í stafrænum heimi.

Hér er aðeins um það að ræða að gera rafræn eintök til að nýta til að vinna úr þeim málfræðilegar og tölfræðilegar upplýsingar sem gagnast við gerð máltæknibúnaðar. Ekki er um það að ræða að verkunum eða hluta þeirra verði dreift á nokkurn hátt. Því verður ekki með nokkru móti séð að samþykkt þessa frumvarps skerði hagsmuni höfunda eða útgefenda.

Nú hafa stjórnvöld ákveðið að hrinda í framkvæmd metnaðarfullri verkáætlun um íslenska máltækni. Flestir eða allir þættir þessarar verkáætlunar byggjast á greiðum aðgangi að miklum mállegum gögnum, og samþykkt þessa frumvarps myndi auðvelda þann aðgang verulega.

Ég hvet eindregið til að frumvarpið verði samþykkt.

Virðingarfyllst,

Eiríkur Rögnvaldsson,
prófessor í íslenskri málfræði
Háskóla Íslands