

Svar

mennta- og menningarmálaráðherra við fyrirspurn frá Kötlu Hólm Þórhildardóttur um máltækni fyrir íslensku.

1. Hyggst ráðherra beita sér fyrir því að heimila undanþágur frá höfundalögum í þágu máltækni til að geta nýtt efni innan kjarnaverkefna máltækninnar sem varin eru með höfundarétti, sbr. skýrsluna *Máltækni fyrir íslensku 2018–2022: verkáætlun?*

Heimild til að nýta höfundavarið efni innan kjarnaverkefna máltækninnar þarf að vera í samræmi við höfundalög. Slík heimild fæst annaðhvort með leyfi höfundarréttarhafa eða á grundvelli undanþágu í höfundalögum. Engin sérstök undanþága er í höfundalögum í dag sem heimilar gagnanám í t.d. orðasöfnum og gagnabönkum vegna máltæknaverkefna. Hins vegar hefur nýlega verið samþykkt tilskipun (ESB) 790/2019 um höfundarétt á hinum stafræna innri markaði sem fellur undir EES-samninginn. Þar er að finna tvö ákvæði um undanþágur vegna gagnanáms (e. text and data mining) sem lögfesta skal í höfundalögum. Annars vegar er um að ræða undanþágu fyrir rannsóknarsamtök og menningarstofnanir til að framkvæma gagnanám í vísindatilgangi í þeim verkum sem viðkomandi samtök eða stofnun hafa löglegan aðgang að. Hins vegar er ákvæði sem heimilar gagnanám í öllum löglega aðgengilegum verkum svo framarlega sem réttthafar slíkra verka hafi ekki gert sérstakan fyrirvara um hið gagnstæða. Verði tilskipunin tekin upp í EES-samninginn er ljóst að Ísland verður skuldbundið til að taka upp slíkar undanþágur vegna gagnanáms í höfundalög.

Samkvæmt samningi ráðuneytisins við Almennaróm skal allur hugbúnaður þróaður innan kjarna- og innviðaverkefna í máltækniáætluninni vera gefinn út með Apache 2.0-leyfi, samþærilegu eða rýmra leyfi. Öll gögn sem nýtt eru í kjarnaverkefnum máltækninnar skulu einnig hafa skilgreind notkunarleyfi. Leyfi vegna nýtingar gagna skal vera CC BY 4.0, sem heimilar margs konar nýtingu efnis, eða rýmra leyfi, ef nokkur kostur er með tilliti til persónuverndarsjónarmiða og hugverkaréttinda þriðja aðila. Apache 2.0-leyfið er mjög opið leyfi sem leyfir ótakmarkaða notkun hugbúnaðar og hugbúnaðarkóða.

Ófrávíkjanleg regla er að gögn sem unnið er með innan máltækniáætlunarinnar séu með skilgreindum notkunarleyfum. Því skulu gögn sem verða til eða eru þróuð innan áætlunarinnar gefin út með opnum alþjóðlegum leyfum, enda er markmið áætlunarinnar að hámarka notkunarmöguleika og nýtingu gagnanna. Leitast skal við eins og mögulegt er að gögn sem aflað er frá þriðja aðila til að nota í máltæknaverkefnum innan áætlunarinnar verði gefin út með opnu leyfi. Þegar því verður ekki komið við skal leitast við að gefa gögnin út með eins litlum takmörkunum og mögulega er unnt. Almenn er þess krafist í flestum leyfum að uppruna eða höfundar gagnanna sé getið.

Um er að ræða kröfur samkvæmt opnum stöðlum og eiga samningar við aðila sem þróa ný kjarna- og innviðaverkefni að endurspeglu þessar kröfur svo verkefnið nýtist sem best.

2. *Hvernig miðar vinnu við kjarnaverkefni verkáætlunarinnar í ljósi þess að fullnægjandi aðgangur að íslensku efni er ekki til staðar?*

Vinnu við kjarnaverkefni verkáætlunar miðar vel. Þetta eru fimm verkefni sem snúa að talgervli, talgreini, þýðingarvél, málrýni og málföngum. Gert er ráð fyrir að hin eiginlega gagnasöfnun og rannsóknarvinna vegna kjarnaverkefnanna verði komin í gang í sumar. Allir aðilar eru sammála um að afar mikilvægt sé að vanda til við þessa vinnu.