



HÁSKÓLI ÍSLANDS

Staða og horfur í íslenskri máltækni

ELRC-vinnustofa

28. september 2018

Eiríkur Rögnvaldsson



Upphaf skipulegs máltæknistarfs á Íslandi



- Nefnd til að kanna stöðu og möguleika tungutækni á Íslandi
 - skipuð af menntamálaráðherra í september 1998
- Skilaði álitum í febrúar 1999
 - lagði til 225-250 m.kr. fjárveitingu á ári í 4-5 ár til að ná sjálfbærni
 - samtals voru veittar 133 m.kr. á árunum 2000-2004
- Ýmis verkefni upprunnin í þessu
 - t.d. [BÍN](#), [MÍM](#) o.fl.



Meginforsendur íslenskrar máltækni

- Hugbúnaður, m.a. til að
 - greina íslensk hljóð, orð, beygingar, setningagerð
 - þýða milli íslensku og annarra tungumála
 - mynda íslensk málhljóð, orð og setningar
- Gagnasöfn, m.a.
 - orðasöfn með beygingum og framburði orða
 - setningafræðilega greind textasöfn
 - orðasöfn með merkingarvenslum
 - söfn af upptökum talaðs máls



Nauðsynlegt er að skapa aðstæður

- Öflugt rannsóknar- og þróunarumhverfi
 - þar sem áhersla verði lögð á nýsköpun
 - byggt upp að verulegu leyti innan háskólanna
 - en í nánum tengslum við fyrirtæki
- Mannauður
 - nauðsynlegt er að mennta fólk í máltækni
 - byggja upp innlenda þekkingu og reynslu
 - með þverfaglegu grunnnámi og meistaranámi
 - í samvinnu ólíkra háskóla og háskóladeilda



Fræðilegt máltæknistarf á Íslandi frá 2001

- Skipulegt máltæknistarf hófst árið 2001
 - á Árnastofnun og við HÍ
 - HR hóf þátttöku árið 2005
 - starfið naut styrkja úr tungutækniáætlun fyrstu árin
- Nokkur verkefni hafa fengið styrki úr sjóðum
 - [*Viabale Language Technology Beyond English*](#)
 - öndvegisstyrkur Rannsóknasjóðs 2009-2011; 43,5 m.kr.
 - [*META-NORD*](#) (norðausturhluti [*META-NET*](#))
 - styrkur ICT-PSP (FP7) 2011-2013; 32 m.kr.



Málföng fyrir íslensku



Þessi vefsíða inniheldur tengla á málgögn, máltól og vefþjónustu á sviði íslenskrar máltækni. Auðvelt er að nálgast þessi málföng sem sum hver eru tilbúin en önnur eru enn í vinnslu. Ef þið vitið um máltækniverkfni sem ætti að vera hér inni sendið þá póst á sigruhel@hi.is.

Málskýrsla

Íslensk tunga á stafrænni öld

Ráðstefna

Máltækni fyrir alla - 27. apríl 2012

Máltól

CombiTagger
IceNLP
Apertium-is-en
hunspell-is

Málheildir - textaskrár

Íslensk orðtíðnibók
Icelandic Parsed Historical Corpus - IcePaHC
Mörkuð íslensk málheild
Risamálheild
Gullstaðall
Íslenskur orðasjóður
Fornritin

Málheildir - texta- og hljóðskrár

Hjal málheildin
Alþingisumræður
Jensson-málheildin
Þór-málheildin
RÚV-málheildin
Málrómur
Alþingisræður til talgreiningar

Mállýsingar

Framburðarorðabókin
Almenn framburðarorðabók fyrir talgreiningu
Mynstur og setningar

Orðasöfn

Beygingarlýsing íslensks nútímamáls
MerkOr - Íslenskur merkingarbrunnur
Íðorðabankinn
ISLEX - íslensk orðabók
Hljóðupptökur ISLEX
Þýðingarmiðstöð utanríkisráðuneytisins
Íslenskt orðanet
IceWordNet
Orðskiptingar

Málföng.is

- Safn íslenskra málfanga (*e. language resources*)
 - komið upp í tengslum við META-NORD
- Sum málföng alveg opin
 - önnur með stöðluðum leyfum (CC, LGPL)
 - eða með sérsömdum leyfum og skilmálum



Úttekt á stöðu íslensku í stafrænum heimi

META-NET

White Paper Series Hvitbókaröð

THE ICELANDIC LANGUAGE IN THE DIGITAL AGE
ÍSLENSK TUNGA Á STAFRÆNNI ÖLD

Eiríkur Rögnvaldsson
Kristín M. Jóhannsdóttir
Sigrún Helgadóttir
Steinþór Steingrímsson

 Springer

- *Íslensk tunga á stafrænni öld*
 - skýrsla META-NET, 2012
- Könnun á stöðu 30 evrópumála
 - gagnvart stafrænni framtíð
- Flokkað eftir stöðu málanna í:
 - talvinnslu
 - vélþýðingum
 - textagreiningu
 - málföngum
- Íslenska í næstneðsta sæti



Stýrihópur og verkáætlun

- Stýrihópur íslenskrar máltækni
 - skipaður af mennta- og menningar-
málaráðherra í október 2016
 - fól sérstakri þriggja manna nefnd að
semja verkáætlun til fimm ára
 - annarri nefnd var falið að gera skýrslu
um kennslu
- Áætlunum skilað í júní 2017
 - *Máltækni fyrir íslensku 2018-2022*
 - [Skýrsla starfshóps um stöðu og fram-
tíð menntunar á sviði máltækni](#)

The image shows the cover of a report titled 'MÁLTÆKNI FYRIR ÍSLENSKU 2018-2022 VERKÁÆTLUN'. The cover has a teal background with a repeating pattern of small text. At the top, there is a list of names and titles, including BERGMÁL, RATLJÓST, and others. Below the title, the authors' names are listed: ANNA BJÖRK NIKULÁSDÓTTIR, JÓN GUÐNASON, and STEINÞÓR STEINGRÍMSSON.



Stjórnarsáttmáli og fjáráukalög

- Stjórnarsáttmáli, nóvember 2017
 - „Fjármögnuð verði aðgerðaáætlun um máltækni þannig að íslensk tunga verði gjaldgeng í stafrænum heimi.“
- Fjáráukalög, desember 2017
 - „Áætlaður heildarkostnaður áætlunarinnar árin 2018–2022 er 2.338 m.kr. og af því er gert ráð fyrir að framlag nýsköpunarfyrirtækja verði 500 m.kr. og að ríkissjóður leggi til 1.838 m.kr. Auk þess er gert ráð fyrir að árlegur rekstrarkostnaður og kostnaður við menntun á sviði máltækni og CLARIN samstarf verði 75 m.kr. á ári eða 375 m.kr. á tímabilinu. Áætlað heildarframlag ríkissjóðs verður því 2.213 m.kr.“



Fjárlög 2018

- Markmið
 - „Byggja upp innviði íslenskrar máltækni í opnum aðgangi þannig að íslenska verði fullgild í stafrænum heimi í samræmi við stjórnarsáttmála. Fyrir liggur verkáætlun um máltækni fyrir íslensku 2018–2022, þar sem lögð er áhersla á uppbyggingu innviða, nýsköpun í máltækni og samstarf og klasamyndun.“
- Fjárveiting
 - „Gert er ráð fyrir 450 m.kr. til máltækniverkefnis [...]. Auk þessa er fyrirhugað að nýta 100 m.kr. af safnlið. Þá er fyrirhugað að Markáætlun komi að fjármögnun verkefnisins.“



Meistaránám í máltækni frá 2001

- Meistaránám í máltækni hófst við HÍ 2001
 - í samvinnu íslensku og tölvunarfræði
 - nemendur teknir inn tvö ár í röð
 - tveir luku námi og tveir héldu áfram annars staðar
- Námið endurvakið 2007
 - í samvinnu íslensku við HÍ og tölvunarfræði við HR
 - nemendur teknir inn annað hvert ár nema 2011
 - 11 hafa lokið námi eða eru að ljúka því
 - ekki teknir inn nýir nemendur 2017



Nýr samningur HÍ og HR um meistaranám

Háskóli Íslands (HÍ) og Háskólinn í Reykjavík (HR) gera með sér eftirfarandi samstarfssamning:

Háskóli Íslands og Háskólinn í Reykjavík hafa með sér samstarf um meistaranám á sviði máltækni (e. *language technology*). Stúdentar innritast í annan hvorn skólann (heimaskóla) og brautskrást þaðan og geta tekið hluta náms síns í samstarfsskólanum og hugsanlega erlendum samstarfsháskólum.

Hvor skóli um sig semur kennsluskrá og námsskipunarreglur og skilgreinir námsefni og námskröfur fyrir þá nemendur sem þar eru skráðir til náms og ber ábyrgð á náminu í heild og veitingu prófgráðu við námslok.

Hvor skóli um sig annast inntöku nemenda í viðkomandi skóla og ábyrgist að þeir sem eru skráðir í meistaranám í máltækni innan Íslensku- og menningardeildar HÍ eða tölvunarfræðideildar HR geti tekið allt að 50 einingar í samstarfsskólanum, án þess að þeir verði krafðir um skrásetningargjöld/skólagjöld. Þessi heimild gildir aðeins fyrir námskeið sem bæði heimaskólinn og samstarfsskólinn telja eðlilegan þátt í máltækninámi í viðkomandi skóla.

Lokaverkefni nemenda er unnið við heimaskóla þeirra undir leiðsögn kennara þar, en í einstökum tilvikum er þó hugsanlegt að semja um að leiðbeinandi komi frá samstarfsskólanum.

Skipuð verður sérstök nefnd með tveimur fulltrúum frá hvorum skóla um sig til að fylgjast með framvindu og framkvæmd samstarfssamnings.

Samstarfssamningur þessi gildir til 30. júní 2023, en skal endurskoðaður að tveimur árum liðnum frá undirritun. Hann er uppsegjanlegur af hálfu hvors skóla um sig með árs fyrirvara.



Úr kennsluskrá HÍ 2018-2019

Máltækni

Máltækni (e. *language technology*) er þverfaglegt rannsóknar- og þróunarsvið sem spannar m.a. tölvunarfræði, málvísindi, gervigreind, tölfræði og sálfræði. Máltækni hefur það að markmiði að próa búnað sem getur unnið með, skilið og myndað mannlegt mál, þannig að það verði nothæft í samskiptum manns og tölvu. Háskóli Íslands og Háskólinn í Reykjavík eiga samstarf um meistaranám í máltækni. Unnt er að fara mismunandi leiðir í máltækninámi á Íslandi.

- Nemendur geta innritast í meistaranám í máltækni við Íslensku- og menningardeild Háskóla Íslands og útskrifast þaðan með M.A.-próf í máltækni, en mega taka allt að 50 einingar við Háskólann í Reykjavík án þess að greiða þar skólagjöld.
- Nemendur geta innritast í meistaranám í máltækni við tölvunarfræðideild Háskólans í Reykjavík og útskrifast þaðan með M.Sc.-próf í máltækni, en mega taka allt að 50 einingar við Háskóla Íslands án þess að greiða þar innritunargjöld.
- Nemendur geta innritast á máltæknikjörsvið innan meistaranáms í tölvunarfræði við Háskóla Íslands og útskrifast þaðan með M.Sc.-próf í tölvunarfræði. Þeir geta tekið námskeið við Íslensku- og menningardeild en ekki námskeið við Háskólann í Reykjavík nema greiða þar skólagjöld.

Sérstakar námsskipunarreglur gilda um hverja þessara námsleiða fyrir sig. Hér er aðeins lýst M.A.-námi í máltækni við Íslensku- og menningardeild. Upplýsingar um inntökuskilyrði og námsskipan er að finna á heimasíðu Íslensku- og menningardeildar.



Úr kennsluskrá HR 2018-2019

Máltækni MSc

Máltækni er rannsóknar- og þróunarsvið sem hefur það að markmiði að þróa búnað sem getur unnið með og skilið náttúruleg tungumál og stuðlað að notkun þeirra í samskiptum manns og tölvu.

Hugvísindi og tækni

Meistaránám í máltækni er tveggja ára, 120 ECTS nám við tölvunarfræðideild. Námið hentar bæði nemendum með BA-próf í hugvísindagreinum og BSc-próf í raungreinum. Markmið með náminu er tvíþætt: annars vegar að útskrifa nemendur með þekkingu til að stjórna verkefnum og útfæra lausnir á sviði máltækni, hins vegar að undirbúa nemendur undir doktorsnám á sviðinu.

Staða íslenskunnar í tæknivæddum heimi

Mikil umræða hefur verið undanfarin misseri um notkun íslensku í samskiptum við hin ýmsu tæki. Bent hefur verið á að bjóða þurfi upp á íslensku í hinum ýmsu viðmótum tækja og er það meðal viðfangsefna máltækni.

Samstarf tveggja háskóla

Meistaránám í máltækni er sameiginlegt nám Háskólans í Reykjavík (HR) og Háskóla Íslands (HÍ). Námið er ætlað bæði fyrir nemendur með BS-próf í tölvunarfræði (eða skyldum greinum) og nemendur með BA-próf í hugvísindagreinum (almennum málvísindum, íslensku og erlendum tungumálum). Þeir sem hafa lokið háskólaprófi í öðrum skyldum greinum geta einnig sótt um inngöngu og er þá undirstaða þeirra metin sérstaklega.

Nemendur þurfa að vera skráðir í meistaránám í máltækni í HR eða HÍ en geta tekið námskeið í báðum skólunum. Nemandi útskrifast frá þeim skóla sem hann er skráður í og vinnur að jafnaði meistaraverkefni undir handleiðslu kennara í viðkomandi skóla.

Nemandi sem skráður er í HR útskrifast með MS-gráðu í máltækni en nemandi sem skráður er í HÍ útskrifast með MA-gráðu í máltækni. Sérstakar námsskipunarreglur gilda um meistaránámið í hvorum skóla fyrir sig og skulu nemendur lúta námsskipunarreglum heimaskóla síns.

Markmið með náminu er tvíþætt: annars vegar að útskrifa nemendur með þekkingu til að stýra verkefnum og útfæra lausnir á sviði máltækni; hins vegar að undirbúa nemendur undir doktorsnám á sviðinu.



Staðan og framtíðin

- Mennta- og menningarmálaráðuneytið hefur samið við sjálfseignarstofnunina Almennaróm
 - um að vera miðstöð máltækniáætlunar
 - og taka að sér framkvæmd verkáætlunarinnar
- Háskólar, rannsóknastofnanir og fyrirtæki þurfa að búa sig undir virka þátttöku í verkefnum
 - í verkefninu eru miklir möguleikar fyrir ýmsar greinar
 - máltækni er þverfagleg og nær til allra fræðasviða
 - ekki bara málvísinda og tölvunarfræði



Markáætlun í tungu og tækni



Markáætlun í tungu og tækni er opin samkeppnissjóður sem starfar samkvæmt lögum um opinberan stuðning við vísindarannsóknir (nr. 3/2003 með áorðnum breytingum [☞]). Áherslur áætlunarinnar eru ákvarðaðar af Vísinda- og tækniráði, en hún er fjármögnuð af mennta- og menningar- málaráðuneyti. Rannís er umsýsluaðili sjóðsins. Markáætlun styður við verkefni í máltækni í því skyni að gera íslensku gjaldgenga í stafrænni upplýsingatækni og stuðla að notkun hennar á þeim vettvangi. Hlutverk Markáætlunarinnar er því annars vegar að vernda og efla íslenska tungu, og hins vegar að auðvelda nýtingu nýrrar

samskiptatækni í íslensku samfélagi, til hagsbótar almenningi, stofnunum og fyrirtækjum.

Markáætlun veitir styrki samkvæmt þeim áherslum sem eru skilgreindar fyrir sjóðinn, og á grundvelli faglegs mats á gæðum verkefna, hversu viðtækt verkefnið er, þörf á afurðum verkefnisins, færni þeirra einstaklinga sem stunda rannsóknirnar og aðstöðu þeirra til að sinna verkefninu. Máltæknisjóður og Markáætlun í tungu og tækni eru keyrð sameiginlega árið 2018.

Auglýst er eftir umsóknum um styrki úr Markáætlun í tungu og tækni fyrir styrkárið 2018-2019. Umsóknarfrestur er til kl. 16:00 þann 9. nóvember 2018.



Máltækni snýst ekki bara um málvernd

- Þróun íslenskrar máltækni er mikilvæg fyrir
 - jafnrétti og mannréttindi
 - tækniþróun og nýsköpun
 - hagkvæmni og hagræðingu í fyrirtækjarekstri
 - skilvirkni í opinberri þjónustu
 - hraða og öryggi í heilbrigðisþjónustu
 - almannavarnir við náttúruvá
- Máltækni snertir öll svið mannlífsins
 - eins og tungumálið sjálft



Takk fyrir áheyrnina!

eirikur@hi.is

<http://uni.hi.is/eirikur>