

Íslensk máltækni í 20 ár

*Að nota íslensku
í tölvum og tækjum*

3. nóvember 2018



VÍSINDAFÉLAG ÍSLENDINGA
1918-2018



HÁSKÓLINN Í REYKJAVÍK
REYKJAVIK UNIVERSITY



Eiríkur Rögnvaldsson

Háskóla Íslands

HÁSKÓLI ÍSLANDS





Máltækniverkefni fyrir 2000

- Tíðnirannsóknir og orðstöðulyklar
 - Baldur Jónsson (1970-1980)
- Vélrænar þýðingar byggðar á venlamálfræði
 - Stefán Briem (1981-2000)
- Vélræn mörkun fyrir *Íslenska orðtíðnibók*
 - Stefán Briem (1985-1990)
- Púki – stafsetningarleiðrétting
 - Friðrik Skúlason (1987-2000)
- Talgervill Öryrkjabandalagsins
 - Pétur Helgason, Höskuldur Þráinsson, Páll Jensson (1989-1990)



Upphaf skipulegs máltæknistarfs

- Nefnd til að kanna stöðu og möguleika tungutækni á Íslandi
 - skipuð af menntamálaráðherra í september 1998
 - formaður Rögnvaldur Ólafsson
- Skilaði álitum í febrúar 1999
 - lagði til 225-250 m.kr. á ári í 4-5 ár
 - samtals voru veittar 133 m.kr.
 - á árunum 2000-2004
- Markar upphaf skipulegs máltæknistarfs á Íslandi





Meginmarkmið

- Meginmarkmið Íslendinga hlýtur að vera að unnt verði að nota íslenska tungu, ritaða með réttum táknum, sem víðast innan tölvu- og fjarskiptataekningarinnar. Þar verður þó að sjálfsgöngu að sníða sér stakk eftir vexti.
- Það er mikið verkefni að gera íslensku gjaldgenga á öllum sviðum, við allar aðstæður. Því verður að leggja megináherslu á þá þætti sem varða daglegt líf og starf alls almennings, eða munu gera það á næstu árum.



Afurðir tungutækniáætlunar

- Textaverkefni
- Beygingarlýsing íslensks nútímamáls
 - Orðabók Háskólans
- Þjálfunarlíkan fyrir málfræðilegan markara
 - Orðabók Háskólans o.fl.
- Mörkuð íslensk málheild
 - Orðabók Háskólans
- Taltækniverkefni
- Hjal – íslensk talgreining
 - Hex, Síminn, HÍ o.fl.
- Ragga – íslenskur talgervill
 - Hex, Síminn, HÍ o.fl.
- Beygingar- og málfræðigreinikerfi
 - Friðrik Skúlason



Máltækniverkefni 2005-2010

- [IceNLP](#)
- IceTagger – markari
– HR – Hrafn Loftsson
- IceParser – þáttari
– HR – Hrafn Loftsson
- Lemmald – lemmari
– Anton Karl Ingason
- [Viable LT beyond English](#)
- MerkOr – merkingargreining
– HÍ – Anna Björk Nikulásdóttir
- IcePaHC – trjábanki
– HÍ – Joel Wallenberg o.fl.
- Apertium – vélþýðing
– HR – Hrafn Loftsson



Máltækniverkefni 2011-2017

- IVONA talgervill
 - Blindrafélagið
- Talgreining í Android-símum
 - Google – HR
- Talgreining fyrir Alþingi og fyrir röntgendeild LSP
 - HR
- Risamálheild
 - Árnastofnun og HÍ
- Gullstaðall
 - Árnastofnun og HÍ
- Skrambi
 - Árnastofnun



Þátttaka í META-NET og afurðir verkefnisins



- [META-NORD](#) (hluti [META-NET](#))
 - ICT-PSP (FP7) 2011-2013; 32 m.kr.
- *Íslensk tunga á stafrænni öld*
 - skýrsla um stöðu máltækni á Íslandi
 - ein af 30 um jafnmörg tungumál
- [Málföng.is](#)
 - safn íslenskra málfanga
 - sem eru ýmist alveg opin
 - eða með stöðluðum leyfum (CC, LGPL)
 - eða sérsömdum leyfum og skilmálum



LREC á Íslandi, maí 2014





Meistaránám í máltækni

- Meistaránám í máltækni hófst við HÍ 2002
 - 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
 - tekið inn annað hvert ár nema 2011 og 2017
 - 11 hafa lokið námi eða eru að ljúka
- Nemendur teknir inn að nýju 2018
 - samkvæmt nýju og breyttu skipulagi



Norræni máltækniháskólinn

- Nordic Graduate School of Language Technology
 - (NGSLT) – 2004-2009
 - kostaður af NordForsk
- HÍ og síðar einnig HR áttu aðild að skólanum
 - nemendur í máltækni námi gátu tekið þar námskeið
 - bæði stutt námskeið og fjarnámskeið með staðlotum
- Eitt námskeið á vegum skólans var á Íslandi
 - „Information Extraction“ í HR vorið 2008



Íslensk málstefna 2009 – meginmarkmið



TILLÖGUR ÍSLENSKRAR MÁLNEFNDAR
AÐ ÍSLENSKRI MÁLSTEFNU

- Að íslensk tunga verði nothæf – og notuð – á öllum þeim sviðum innan tölvu- og upplýsingatækninnar sem varða daglegt líf alls almennings
 - viðmót algengs hugbúnaðar þarf að vera íslenskt
 - til þarf að vera ýmiss konar hugbúnaður sem liðsinnir og leiðbeinir notendum við notkun íslensks máls
 - unnt á að vera að nota íslensku sem samskiptamál við ýmiss konar tölvu- og tæknibúnað



Tvær nefndir

- Nefnd til að gera tillögur um framkvæmd íslenskrar málstefnu hvað varðar íslensku í tölvuheiminum
 - skipuð af menntamálaráðherra í janúar 2010
- Skilaði álitum í júní 2012
 - tillögum nefndarinnar var ekki hrint í framkvæmd
- Nefnd um notkun íslensku í stafrænni upplýsingatækni
 - skipuð af menntamálaráðherra í september 2014
 - í kjölfar þingsályktunar vorið 2014
- Skilaði álitum í des. 2014
 - tillögum nefndarinnar var ekki hrint í framkvæmd



Stýrihópur og verkáætlun

- Stýrihópur íslenskrar máltækni
 - skipaður af menntamálaráðherra 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 framtíð menntunar á sviði máltækni*

BERGMÁL RATLJÓST
0111 010101 HUGFANGINN SINDRANDI SAKNA 010101110101010101010111
011010101 VELKOMIN HLJÓÐ HUGHRIF GLUGGAVEDUR 010 010101110101010111010
MAMMA FAGURKERI KÆRLEIKUR BIRTA KOTROSKIN ALÚÐ
011110 BÁRUJÁRN UMHYGGJA ANDVARI FLAUELSMJÚKT LJÓSMÓÐIR 01010110 010101001010
EINSTÖK MJÖLL DÚNMJÚKUR SÓLARLAG DÁSAMLEGT KRADAK 010101110110 010101
BRYNJA ALGLEYMI FOSS HUNDSLAPPADRÍFA ÁST
010101101 H.JARTA BLÆBRIGÐI DINDILL ÖLDUGJÁLFUR BLIKUR VÍÐSÝNI 010101110101010
FIDRINGUR HIMINLIFANDI DALALÆDA SKAFRENNINGUR EINURÐ
1110110101 LJÓSMÓÐIR RÖKKUR LITSKRÚÐUG HARÐJAXL SEIGLA EDRÚ
01010111010101 GULL SJÓNAUKI KJÓLUR JÆJA ÖGURSTUND BLÆR 01010010 01010111011
NÚNA MJÖLL HRYNJANDI UMHYGGJA DÁSAMLEGT 0101011101 010101110
VÍÐSÝNI KÆRLEIKUR HLJÓÐ ANDVARI ALGLEYMI
010101110101 ÍVAF ÖTTA UGLA KOTROSKIN MAMMA FOSS 010101110 010101110101010
010101011101010101110101 AGNARÖGN SKÚMASKOT 01010111010101010101011101
SPÉKOPPAR

**MÁLTÆKNI FYRIR
ÍSLENSKU 2018-2022**
VERKÁÆTLUN

ANNA BJÖRK NIKULÁSDÓTTIR
JÓN GUÐNASON
STEINÞÓR STEINGRÍMSSON



Verkefni sem er lokið eða standa vel

- Helstu tölvuforrit á almennum markaði verði á íslensku (Windows, Word, Excel, Netscape, Internet Explorer, Eudora, ...)
- Unnt verði að nota íslenska bókstafi (áéíóúýðþæöÁÉÍÓÚÝÐÞÆÖ) við allar aðstæður; í tölvum, GSM-símum, textavarpi og öðrum tækjum sem almenningur notar.
- Ákveðnum aðilum (stofnunum eða fyrirtækjum) verði falin ábyrgð á einstökum verkefnum.
- Komið verði upp stórri tölvutækri textaheild með íslenskum textum af sem fjölbreyttustum toga til að byggja áframhaldandi vinnu á.
- Til verði góður íslenskur talgervill sem geti lesið upp íslenskan texta með skýrum og auðskiljanlegum framburði og eðlilegu tónfalli og sem sé skiljanlegur án þjálfunar.



Verkefni sem er ólokið eða bíða

- Unnið verði að þróun málgreiningar fyrir íslensku, með það að markmiði að geta greint íslenskan texta í orðflokka og setningarliði.
- Til verði góð hjálparforrit við ritun texta á íslensku, s.s. orðskipti-forrit, stafsetningarleiðréttingarforrit, málfarsleiðréttingarforrit o.fl.
- Unnið verði að þróun talgreiningar fyrir íslensku, með það að markmiði að til verði forrit sem geti túlkað eðlilegt íslenskt tal.
- Unnið verði að þróun forrita til vélrænna þýðinga milli íslensku og annarra tungumála, m.a. til að auðvelda leit í gagnabönkum.
- Komið verði upp fullgreindu orðasafni (með málfræðilegri og merkingarlegri greiningu) til nota í áframhaldandi vinnu.



Yfirlit

- Tungutækniáætlunin 2000-2004 gerði mikið gagn
 - en henni hefði þurft að fylgja meira fjármagn
 - og öflug vitundarvakning meðal fólks og fyrirtækja
 - án óraunhæfra krafna um sjálfbærni á stuttum tíma
- Ýmislegt hefur þó áunnist á 20 árum
 - sem leggur góðan grunn að væntanlegu stórataki
- Við erum rétt að byrja!



Takk fyrir áheyrnina!

eirikur@hi.is

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