

META  NET

META  NORD

META-NORD og META-NET

Brýr milli tungumála

Eiríkur Rögnvaldsson

Yfirlit

- META-NORD er verkefni sem Norðurlönd og Eystrasaltslönd standa að. Það er eitt þriggja systurverkefna sem hafa að markmiði að efla málleg gagnasöfn sem nýst geti í margvíslegum máltækniverkefnum
- Megintilgangur verkefnanna er að skapa tækni-
legar forsendur fyrir margmála upplýsingasamfélagi í Evrópu þar sem allir geti notað móðurmál sitt við öflun og úrvinnslu upplýsinga

T4ME

- Upphaf: Technologies for the Multilingual European Information Society – [T4ME](#)
 - verkefni styrkt af 7. rammaáætlun ESB
 - þátttakendur frá 13 Evrópulöndum
- T4ME hefur komið á fót [META-NET](#)
 - A Network of Excellence forging the **Multilingual Europe Technology Alliance**
 - META-NET is a Network of Excellence
 - dedicated to building the technological foundations of a multilingual European information society

META-NORD

- Evrópusambandið vildi víkka META-NET út
 - þannig að það næði til allra tungumála EES
- Því voru stofnaðir þrír samstarfshópar
 - í Norður-, Austur- og Suður-Evrópu
- Evrópusambandið fól [Tilde](#) í Riga
 - fyrirtæki með mikla reynslu í máltækniverkefnum
 - að mynda samstarfshóp Norður- og Eystrasaltslanda
- Úr því varð verkefnið [META-NORD](#)

Markmið ESB með verkefnum

- The overall objective is to ease and speed up the provision of online services centred around computer-based translation and cross-lingual information access and delivery
 - The focus is on assembling, linking across languages, and making widely available the basic language resources (models, tools and datasets) used by developers, professionals and researchers to build specific products and applications
 - Emphasis is placed on less-resourced languages

Þátttakendur í META-NORD

- Þátttakendur auk Tilde eru háskólar
 - í Kaupmannahöfn, Bergen, Gautaborg, Helsinki, Tartu, Vilnius og Reykjavík
- Íslenski þátttakandinn er [Máltæknisetur](#)
 - samstarfsvettvangur Málvísindastofnunar HÍ, tölvunarfræðideildar HR og Árnastofnunar
- Fé til verkefnisins er u.þ.b. 365 milljónir króna
 - þar af tæpar 33 milljónir (á gengi dagsins) til Íslands

Norræn upplýsingasetur í máltækni

- Verkefnið er á vissan hátt framhald á starfi norrænna upplýsingasetra í máltækni
 - The aim of these centres is to ensure that future and, as far as possible, existing research results, corpora, tools etc. in the field of language technology, are available and re-usable
- Setrin unnu saman í [NorDokNet](#), 2001-2005
 - þ. á m. íslenskt upplýsingasetur, tungutaekni.is

Íslenski hlutinn

- Íslenski hluti verkefnisins er hafinn
 - þrír starfsmenn hafa verið ráðnir
 - vinnuaðstaða hefur fengist hjá SÁ á Neshaga 16
- Meginverkefni:
 - samning málskýrslu á íslensku og ensku
 - val og söfnun gagna og viðræður við réttihafa
 - lýsing, uppfærsla, aðlögun og frágangur gagna
 - athugun og frágangur á réttindamálum
 - vitundarvakning, kynning, útbreiðsla, áframhald

Málskýrslur (language reports)

- 30-40 síðna [skýrsla](#) fyrir hvert mál
 - á viðkomandi máli og ensku
 - liggja fyrir í lok maí 2011
- Í skýrslunum verður gerð grein fyrir
 - málsamfélaginu og hlutverki málsins í því
 - máltæknirannsóknnum og máltækniiðnaði í landinu
 - hlutverki máltækni (afurða og þjónustu) í landinu
 - lagalegum atriðum, s.s. höfundarréttarmálum

Söfnun og frágangur málgagna

- Val og söfnun málgagna
 - viðræður og samningar við eigendur og rétthafa
- Uppfærsla og aðlögun málgagna
 - ýmiss konar frágangur og endurbætur
 - gögnunum lýst á staðlaðan hátt
 - gengið frá gögnunum á stöðluðu formi
- Gögnin verða gerð aðgengileg á [META-SHARE](#)
 - með stöðluðum leyfum s.s. [Creative Commons](#)

Tegundir málgagna

- Áhersla lögð á þrjár tegundir málgagna:
 - orðanet
 - venslaflokkuð orðasöfn
 - trjábanka
 - setningagreindar málheildir
 - (íð)orðasöfn
 - einmála og margmála
- Þessi gögn á að tengja milli mála eftir föngum
 - til að þau nýtist sem best í margmála verkefnum

Stefna íslenskra stjórnvalda

- „Stefnt skal að því að hugbúnaður sem smíðaður er og fjármagnaður af opinberum aðilum, m.a. í rannsóknar- og þróunarverkefnum, verði endurnýtanlegur. Liður í því er að hugbúnaðurinn sé frjáls.“
 - [Frjáls og opinn hugbúnaður. Stefna stjórnvalda](#)
- Frjálsum og opnum hugbúnaði er yfirleitt dreift með stöðluðum nýtingarleyfum
 - s.s. [GPL](#), [LGPL](#) o.fl.

Íslensk málstefna

- Til að ná því markmiði „að íslensk tunga verði nothæf – og notuð – á öllum þeim sviðum innan tölvu- og upplýsingatækni sem varða daglegt líf alls almennings“ er lagt til:
- „Að málleg gagnasöfn og hugbúnaður til að vinna með íslenskt mál verði gerð opin og frjáls eftir því sem kostur er (sbr. stefnu stjórnvalda um frjálsan og opinn hugbúnað [...])“
 - [Íslensk málstefna](#), samþykkt á Alþingi 12.3. 2009

Að lokum

- Íslensk málgögn eru flest gerð fyrir opinbert fé
 - sjaldnast komið upp í gróðaskyni
 - yfirleitt ekki mikil söluvara
- Það er allra hagur að gera þau aðgengileg
 - við fáum aðgang að gögnum annarra á mót
 - rannsóknir á íslensku máli stóraukast
 - þróun hugbúnaðar og málgagna fyrir íslensku vex
- Deilum gögnum!