



UiT Noregs arktiske universitet

Open Language

Barrierar for å bruka språkteknologi

Sjur Nørstebø Moshagen

Divvun-gruppa, UiT

The Divvun logo, featuring a green checkmark to the left of the text 'JÁBČ' and 'Divvun' below it. The background is a photograph of a snowy, hilly landscape under a blue sky.

JÁBČ
Divvun

Aslak Sira Myhre, Trondheim, sep 2024

- Den store slagsida er ikkje rasisme, kjønn m.m., det er engelsk
- Noreg har ressursane (NB), samfunnsmodellen (samarbeid, rettigheter), og reknekrafta (Sigma, Lumi)
- Vi kan laga modellar for det norske samfunnet (norsk og samisk)

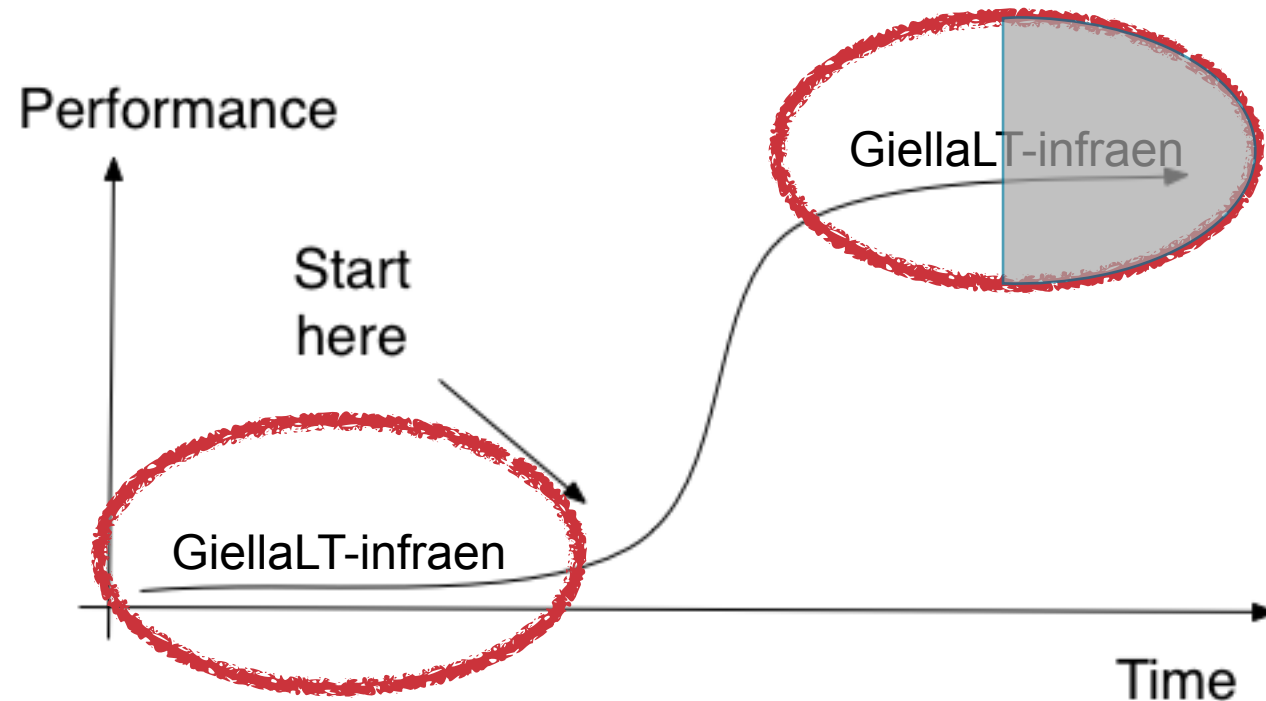


Helsing til Aslak Sira Myhre

- Gode poeng
- MEN:
 - Eit digitalt samfunn treng meir enn KI (t.d. tastatur og retteprogram)
 - (God) språkteknologi er berre halve jobben
 - Den engelske LLM-slagsida har ein parallell:
 - All digitalteknologi har ei språkleg innpakking, og ho er styrt av store, engelskspråklege firma - datateknologien er engelsk

Tre faser i utviklinga

- Dørstokkmila
- Språkarbeidet
- Finpuss og levering
— den siste mila



Hinder

- Tastatur for nettbrett og Android-telefonar
- Retteprogram i nettbaserte Office-pakker
- Språkval (eller mangel på slike):
 - Google Docs
 - MS Office for Mac \Rightarrow kan ikkje lesa opp samisk tekst!
- Taleteknologi og OCR
- Digitale assistentar — helse og heim
- Lokalisering & omsetjing av program og system

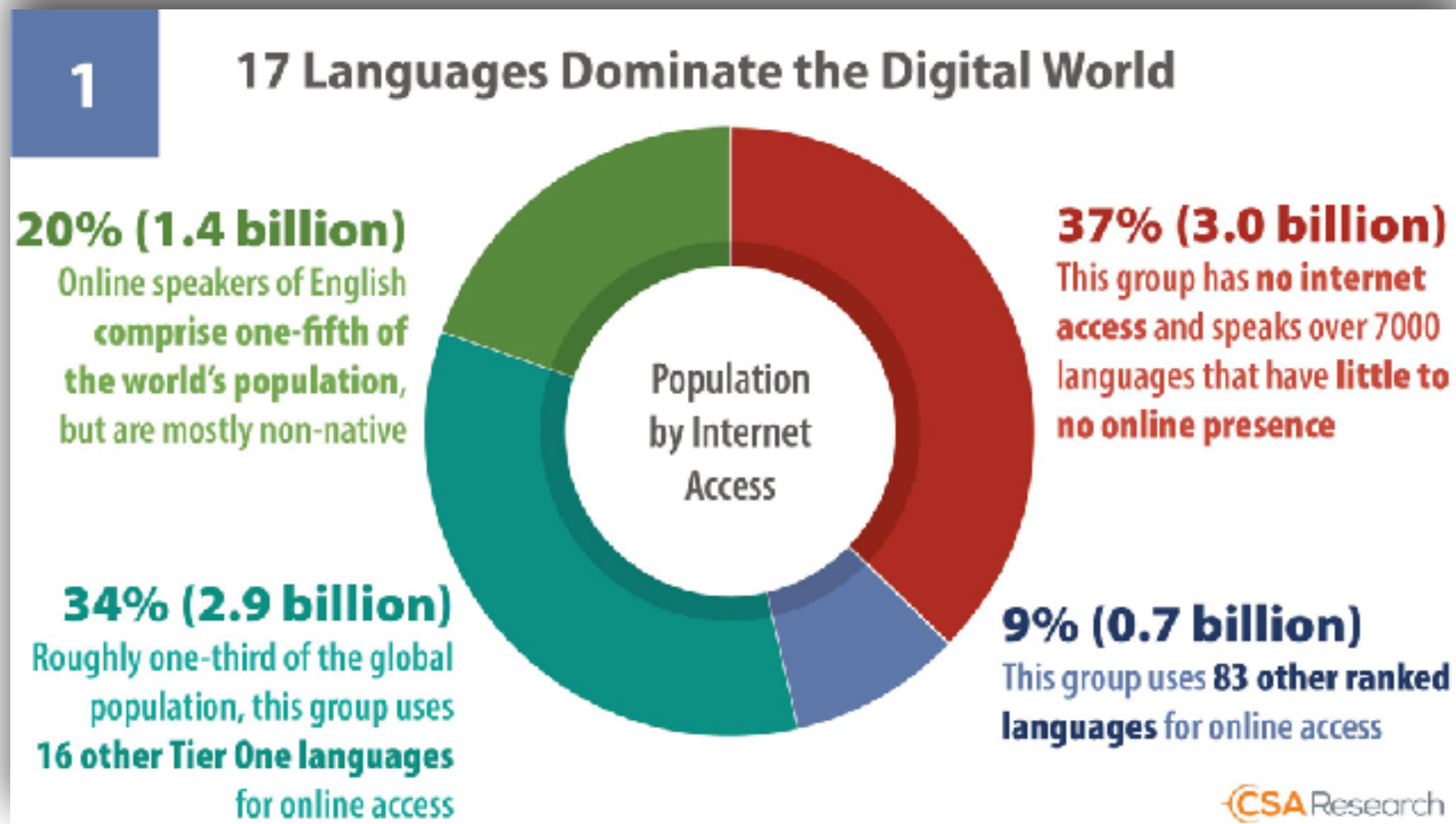
Industriinitiativ

- Finst nokre prosjekt på å laga tastatur for urfolk og minoritetar
 - Google
 - Motorola
 - Apple
 - MS
 - ... m.fl.
- Men det monnar ikkje:
 - 7000+ språk vs 800+ tastatur i GBoard - 10 % tastatur
 - Og kva med alle andre verktøy?

Digital infra = språkleg infra

- Grunnleggjande for samfunnet
- Samfunnet har svært lite styring
 - Økonomi - innkjøp
 - Men: fyrst investera i Divvun-gruppa, og deretter betala til industrien for å gjenskapa Divvun-arbeidet i firma X sitt system?
- Nordisk språkdeklarasjon:
 - Alle skal vera med, men vi har ikkje dei politiske verktøya for å sikra dette

Gjeld mange - 3 mrd



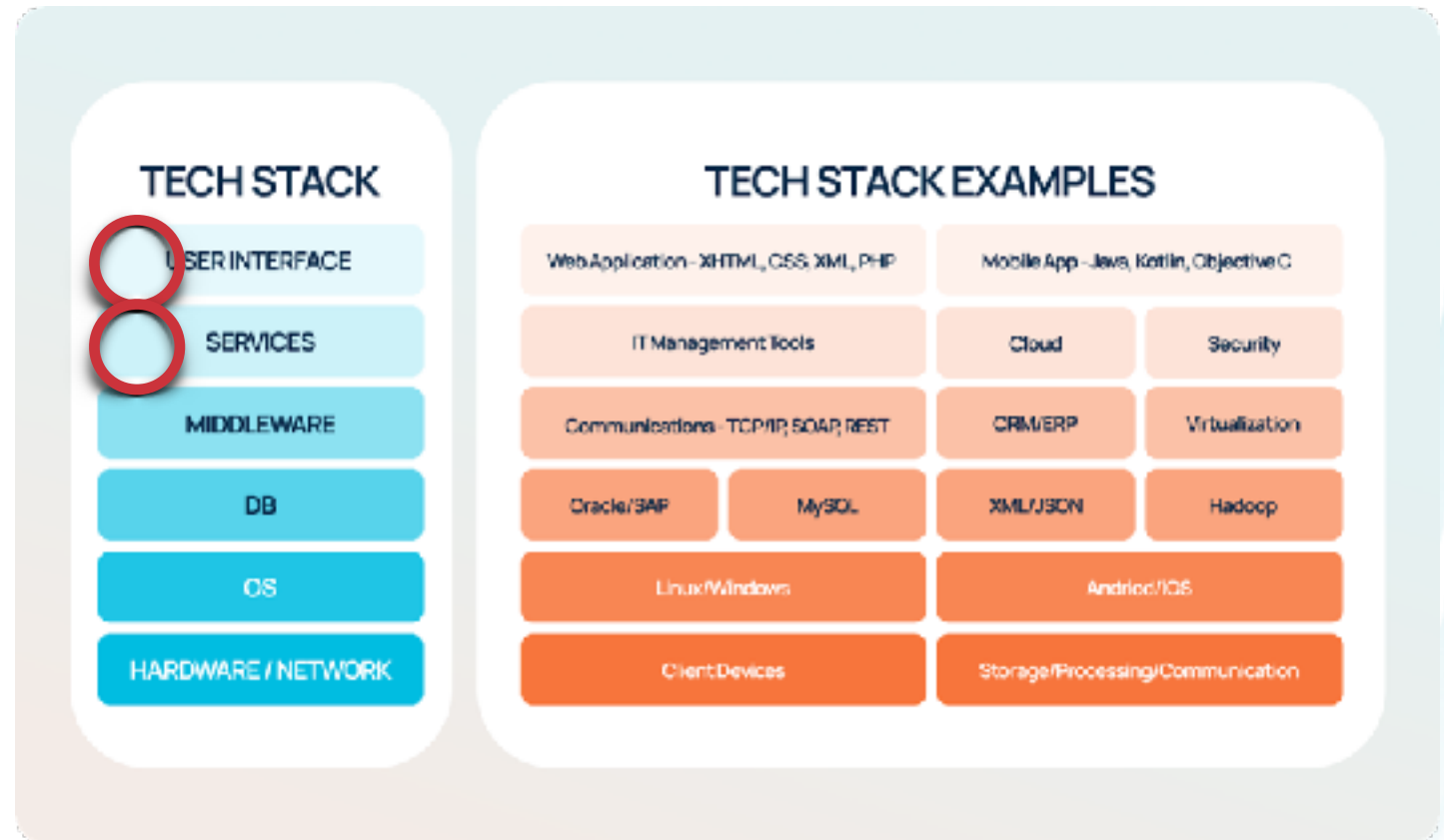
Kjelde: Arle Lommel, https://www.linkedin.com/posts/arlelommel_digitaldivide-languageaccess-inclusivity-activity-7188624362231545858-U5gg

Industrien missar poenget

- Språksamfunna får teknologi på nåde
- Dei har nærmast null styring sjølv
- Eigenutvikla teknologi og verktøy kjem ikkje ut
- Byggjer ikkje kompetanse og språkteknologisk «sjølvberging»
- Vi som samfunn har ikkje styring med det mest sentrale kittet i samfunnsbygget — nemleg språket
 - Ikkje normeringa
 - Ikkje kor vi kan eller får bruka det
 - Bilen pratar til deg: på engelsk, i beste fall på riksmål

Svar: Open Language

- Inspirert av open kildekode
- ... men no berre språkdelen av eit system:
 - Lokaliserbar tekst
 - API-ar for språktenester
- Skal vera open for alle språk, utan spørsmål eller krav
- Enkelt å utvikla for, enkelt å ta i bruk



Kjelde: Apxor: What is a Tech Stack. Raude sirklar lagt til av forfatternen.
(<https://www.apxor.com/blog/tech-stack-for-consumer-apps>)

Open Language — vinn-vinn

- Det meste finst der:
 - Språk og teknologi allereie skilde frå kvarandre
 - Ferdig distribusjonsapparat (App Store, Play Store, mm)
- Ingen nye risikovektorar — etablerte metodar og system ok
- Treng ikkje opna kjeldekode eller forretningshemmelegheiter
- 3 milliardar nye kundar
- Språksamfunna får styring over språka sine igjen

Politisk press

- Endringa kjem ikkje av seg sjølv
- Nordisk samarbeid viktig
 - Mot EU
 - Mot FN i samband med urfolksspråkstiåret?
- Engelsk slagside:
 - Nordiske LLM-ar er bra, men ikkje nok
 - Open Language er grunnleggjande for å hindra at digitalisering fører til språkbyte til engelsk eller eit anna majoritetsspråk



UiT Noregs arktiske universitet

Giitu! Takk!

Sjur Nørstebø Moshagen

divvun.no, giellalt.github.io

✓ JÁBČ
Divvun