

# Introduktion til Sprogteknologi

Forår 2020  
Lecture 1

# Hvem & Hvornår

Tino Didriksen <tino@oqaasileriffik.gl>, 1. del:

- Uge 6: mandag 3/2 + onsdag 5/2 + fredag 7/2
- Uge 7: mandag 10/2 + onsdag 12/2 + torsdag 13/2

Trond Trosterud <trond.trosterud@uit.no>, 2. del:

- Uge 17: mandag 20/4 + onsdag 22/4 + fredag 24/4

Per Langgård <per@oqaasileriffik.gl>, 3. del:

- Uge 22: mandag 25/5 + onsdag 27/5 + torsdag 28/5 + fredag 29/5

# 1. del – Brug af terminalværktøj

Unix, Posix, Linux, BSD - tilsammen kaldet \*nix - har en samling af nyttige terminalværktøj som danner grundlag for alt overbygning.

Formålet med 1. del af kurset er at få jer fortrolige med at arbejde i en terminal, og med de eksisterende analyseværktøj der findes for Kalaallisut.

# Efter disse 2 uger skal i kunne...

1/2

- ls, cd, pwd, mv, rm, mkdir, grep, egrep, tr, sed, sort, uniq, awk, echo, cat, uconv, more, less, diff ... og mange flere
- Pipe data fra et program til det næste med `|`, eller fra en fil med `cat`, eller til en fil med `>`, eller alle tre på samme tid
- Benytte svn (Subversion) versionskontrol
- Bygge Kalaallisut-analysatoren, og genoprette den hvis den går i stykker

# Efter disse 2 uger skal i kunne...

2/2

- Læse og forstå maskin-analyserede tekster
- Manuelt disambiguere (entydigggøre) maskin-analyserede tekster
- Forstå hvorfor det er svært for en maskine at analysere tekster
- Benytte regression testing værktøj
- De vigtigste dele af Regular Expressions

## Som munder ud i at kunne

- Analysere og sammenligne korpora og tekster

# Dagens mål

- At få installeret terminal og Kalaallisut-analysatoren på alle jeres laptops
- At lære jer hvordan i kan lave en frekvensanalyse på ordformer af vilkårlige tekster

# De vigtigste kommandoer

1/2

- **ls** List - viser indholdet af den mappe man står i
- **cd** Change Directory - skifter mappe
- **pwd** Print Working Directory - skriver nuværende mappe til terminalen
- **mv** Move - flytter filer og mapper
- **cat** Concatenate - læser filer og outputter dem til terminalen
- **grep** Fra- eller tilvælger linier i inputtet
- **tr** Translate - ændrer tegn i inputtet
- **sort** Sorterer linier i alfabetisk eller numerisk orden
- **uniq** Unique - springer over linier hvis foregående linie var identisk
- **more** En såkaldt "pager" som viser inputtet på terminalen opdelt i sider

# De vigtigste kommandoer

2/2

- **wc** Word Count - skriver antal linier, ord, og bytes til terminalen
- **rm** Remove - fjerner filer og mapper
- **mkdir** Make Directory - oprettet mapper
- **echo** Skriver en kort besked til terminalen
- **less** Også en "pager", men denne kan søge i det man ser
- **diff** Difference - viser forskellene på 2 filer
- **history** Skriver tidligere kørte kommander til terminalen



# Hjælp i terminalen

Hvis i ikke ved hvordan et program eller kommando fungerer er der flere måder at finde ud af det. E.g., hvis kommandoen er **grep**, så prøv:

- **grep --help**
- **grep -h**
- **grep -?**
- **man grep**

Ikke alle kommandoer har **--help** eller **-h** eller **-?** eller en **man**-page, så man skal prøve sig frem.

# Terminologi

- Mappe, folder, directory, katalog - alle det samme begreb