

Status Quo efter lektion 4

Startede med at kigge på en mulig løsning til opgaven fra sidste gang. Ud fra en ren numerisk statistisk analyse af corpora kan man se at:

- Ataqinartuaraq, UkiutTrettenit, og Facebook benytter @CL-CIT på samme måde - ca. 75% til venstre, 25% til højre.
- Aviscorpus er klart forskellig med 50/50 split.
- Facebook er også klart forskellig med at man næsten slet ikke benytter @CL-CIT.

Hvorfor det forholder sig sådan *is left as an exercise for the reader*. I dette tilfælde kan man dog nok udtænke gode grunde uden at skulle kigge på faktiske eksempler. Men metoden kan bruges til at undersøge mange stilistiske forskelle, og til at be- eller afkræfte formodninger om sproglig brug.

Derefter kørte alle Regression Test.

- Kør alle test-corpora gennem analysen

```
cd ~/langtech/regression/  
./analyse.pl
```

- ... og derefter se om der var nogen uforventede ændringer

```
./compare.pl
```

- ... åben den næst-nederste URL der kommer frem (oftest <http://localhost:3000/>) i en browser, og når man er færdig med den tryk CTRL-C i terminalen.

- Og så startede alle en opdatering af Kalaallisut-analysatoren.

```
time ~/lt-course/lecture01/scripts/update-svn.sh
```

Programmet `time` gør hverken fra eller til - den viser bare hvor lang tid en kommando tager at køre. I VirtualBox tager en opdatering worst-case ca. 45 minutter. På serveren i Canada og på native macOS tager det worst-case ca. 20 minutter.

Hvis dem der har Mac's hellere vil have det hele kørende native så i ikke behøver VirtualBox, så kan i gøre det ved at:

- Installere XCode fra App Store
- I en macOS Terminal (Applications -> Utilities -> Terminal ... eller på dansk Programmer -> Hjælpeprogrammer -> Terminal), køre `xcode-select --install` og vente på at den bliver færdig.
- I en Terminal, køre `sudo xcodebuild -license`
- Installere MacPorts fra <https://www.macports.org/install.php>
- I en Terminal, køre
 - o `cd`
 - o `git clone https://github.com/Oqaasileriffik/lt-course`
 - o `~/lt-course/lecture01/scripts/install-macos.sh`

Hele processen tager mange timer og meget diskplads, hvilket er grunden til at vi ikke gjorde det på den måde første dag. Men hvis i vil arbejde med dette fremover, så skal det være native før eller senere.

Og så gennemgik vi de første stykker Regular Expression. Her er alt dækket af slides og links derfra.

Opgave til onsdag d. 12

- Gå gennem <https://regexone.com/> eller tilsvarende tutorial.
- Leg med regexes i <https://regex101.com/> enten med egne tekster eller med de udelte corpora.

Noter & Hints

Noget jeg bliver ved med at glemme: Kommandoen `history` giver en liste af tidligere kørte kommandoer, men kun hvis man har lukket terminalen på en måde hvor den gemmer historik. Dette gøres ved kommandoen `exit` eller ved at på en tom terminal trykke CTRL-D.

Det er også muligt at konfigurere sådan at historik altid bliver gemt - sådan er det gjort på serveren i Canada. Her er der også en extra stor fil der husker mere end `history`, nemlig filen `~/.history_long` (ja, filnavnet starter med `.`).