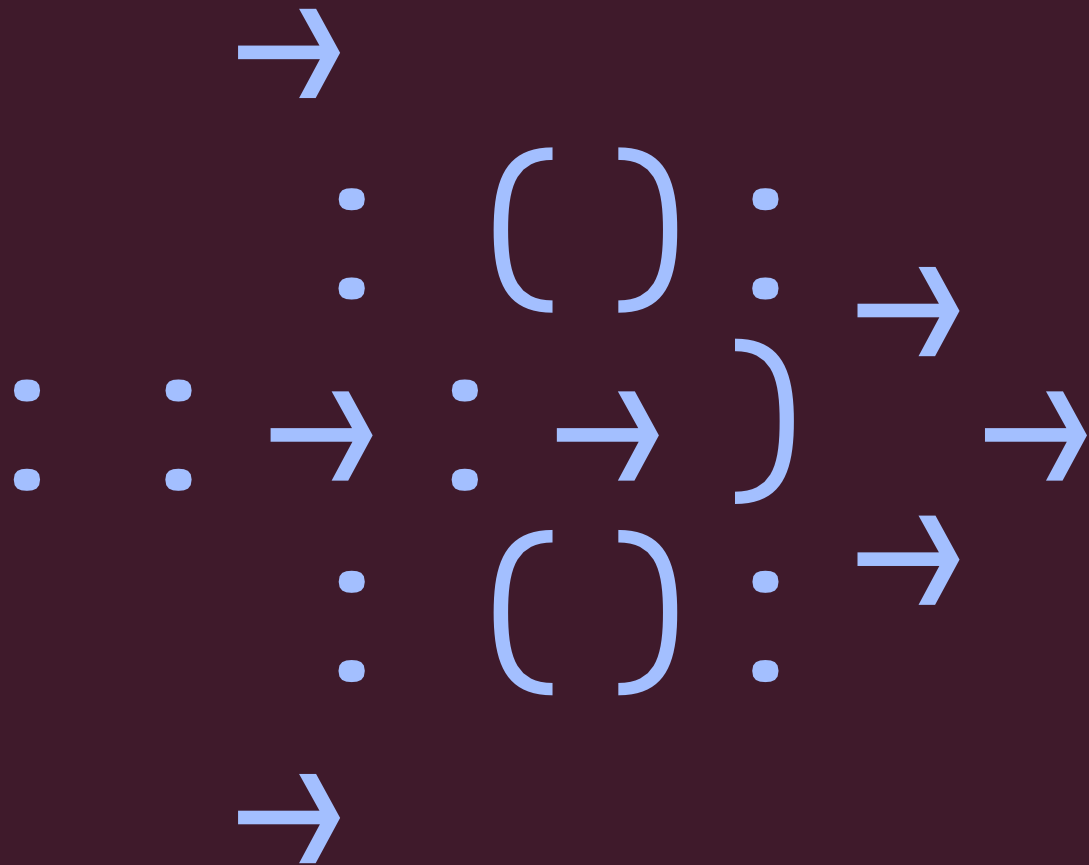


CE

20. September 2024

Workshop om evaluering og benchmarking af sprogmodeller på dansk



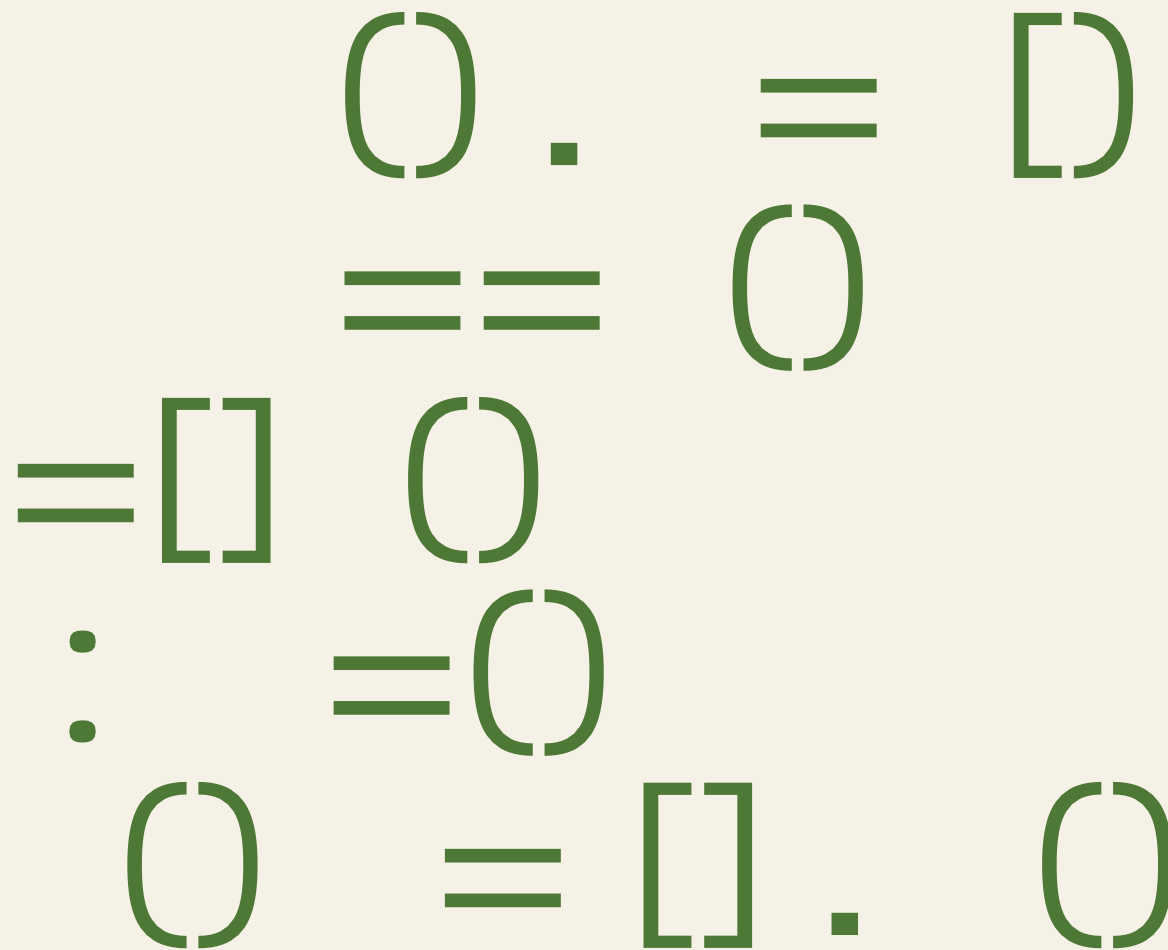


Formål med dagen

1. At gøre status på arbejdet med at evaluere sprogmodeller på dansk.
2. At afdække behovet for yderligere indsatser på området.
3. At udarbejde et udkast til en backlog af opgaver, som Digitaliseringsstyrelsen og andre kan kaste sig over.

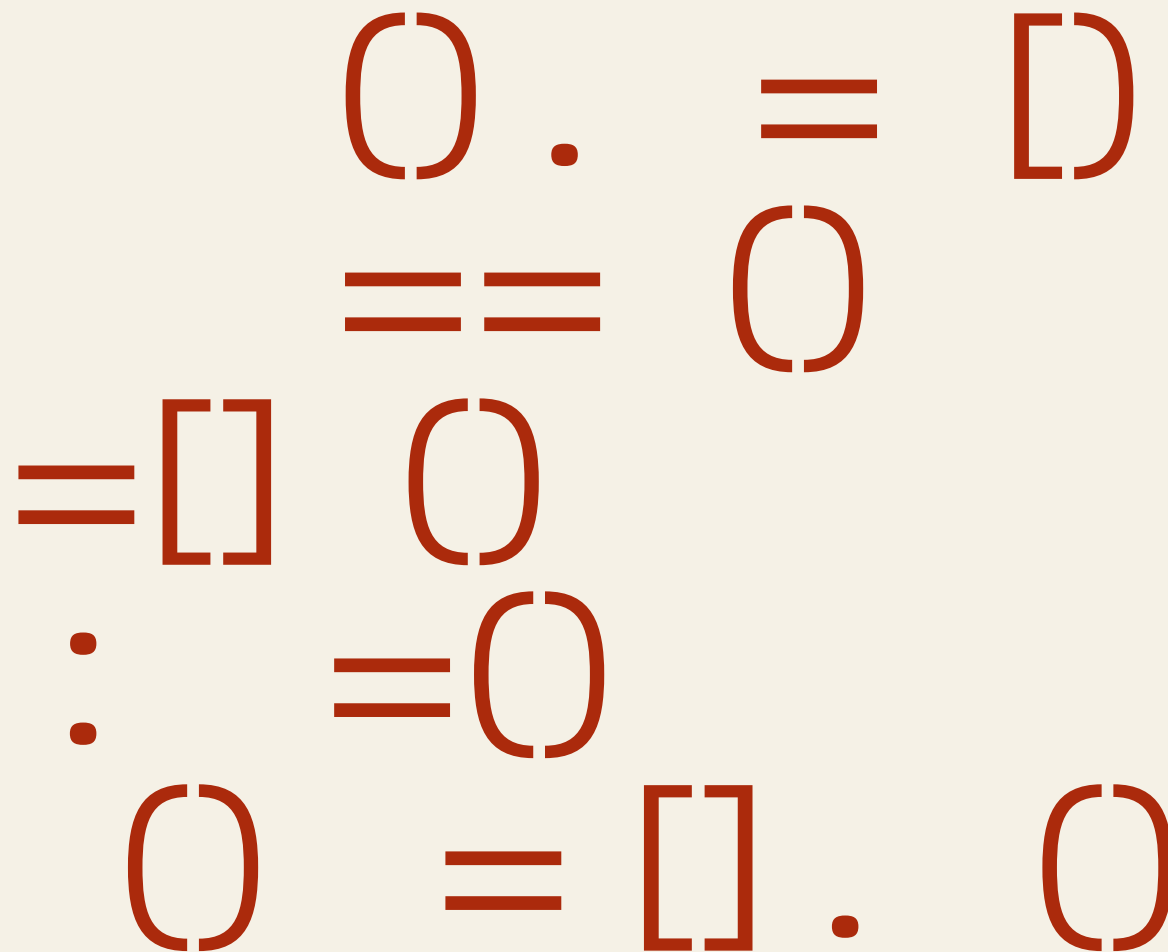
Agenda – Del 1

- Intro til evaluering af sprogmodeller og ScandEval
Dan Saatrup Nielsen (Alexandra Instituttet)
- Scandinavian Embedding Benchmark
Kenneth Enevoldsen (AU)
- Pause
- Danish Semantic Reasoning Benchmark
Bolette Sandford Pedersen (CST) og Nathalie Hau Sørensen (DSL)
- Danoliterate
Søren Vejlgård Holm (DTU)
- Øvrige projekter:
 - Norske projekter ved Hans Christian Farsethås (UIO)
 - 1000 danske talemåder ved Nathalie Hau Sørensen (DSL)



Agenda – Del 2

- Frokost
- Opgave 1: Kortlægning
Hvad har vi nu? Og hvad mangler vi?
- Pause
- Opgave 2: Backlog og prioritering
Hvordan kan vi imødekomme manglerne? Hvem skal involveres i dette arbejde? Og under hvilke forudsætninger?
- Afrunding (ca. 15:00)
- Mulighed for networking
Vi har lokalerne frem til kl. 16:00.



💡 Missing pieces for Danish NLP

Vote and suggest missing pieces for Danish NLP

↑ 1



Instruction tuning dataset for Danish

[KennethEnevoldsen](#) started on Jan 8 in [Missing pieces for Danish NLP](#)



1

↑ 1



Distilled model for high-throughput Danish NLP model

[help wanted](#) [model](#)

[KennethEnevoldsen](#) started on Dec 8, 2023 in [Missing pieces for Danish NLP](#)

↑ 1



Bias Benchmark for Danish language models

[help wanted](#) [evaluation](#)

[KennethEnevoldsen](#) started on Dec 8, 2023 in [Missing pieces for Danish NLP](#)

↑ 3



Document encoders for Danish

[in-progress](#) [model](#)

[KennethEnevoldsen](#) started on Dec 8, 2023 in [Missing pieces for Danish NLP](#)

↑ 3



Generative Language Models for Danish

[in-progress](#) [model](#)

[KennethEnevoldsen](#) started on Dec 8, 2023 in [Missing pieces for Danish NLP](#)

↑ 2



Benchmark for generative models for Danish

[in-progress](#) [evaluation](#)

[KennethEnevoldsen](#) started on Dec 8, 2023 in [Missing pieces for Danish NLP](#)

↑ 2



Benchmarks for Danish document embeddings

[in-progress](#) [evaluation](#)

[KennethEnevoldsen](#) started on Dec 8, 2023 in [Missing pieces for Danish NLP](#)



0

Benchmarks for Danish document embeddings #196

[KennethEnevoldsen](#) started this conversation in [Missing pieces for Danish NLP](#)



[KennethEnevoldsen](#) on Dec 8, 2023 [Maintainer](#)

...

Statement of need

Multiple new and old tools seeks improve the representation of documents to better retrieve documents for search or for generative systems such as RAGs. It is currently unknown how well models perform on Danish for this specific task. Constructing a reasonable benchmark allow for meaningful development and selection of document encoders for Danish (be it multilingual or not).

Current state

A benchmark is currently being constructed at:

<https://github.com/KennethEnevoldsen/scandinavian-embedding-benchmark>

How to contribute:

If you wish to contribute to this task please see the associated GitHub for the project.

↑ 2



Dan Saattrup Nielsen

Intro til evaluering af sprogmodeller og ScandEval



Kenneth Enevoldsen

Scandinavian Embedding Benchmark



Bolette Sandford Pedersen og Nathalie Hau Sørensen

Danish Semantic Reasoning Benchmark



Søren Vejlgård Holm

Danoliterate



Hans Christian Farsethås

Norske projekter



Nathalie Hau Sørensen

1000 danske talemåder

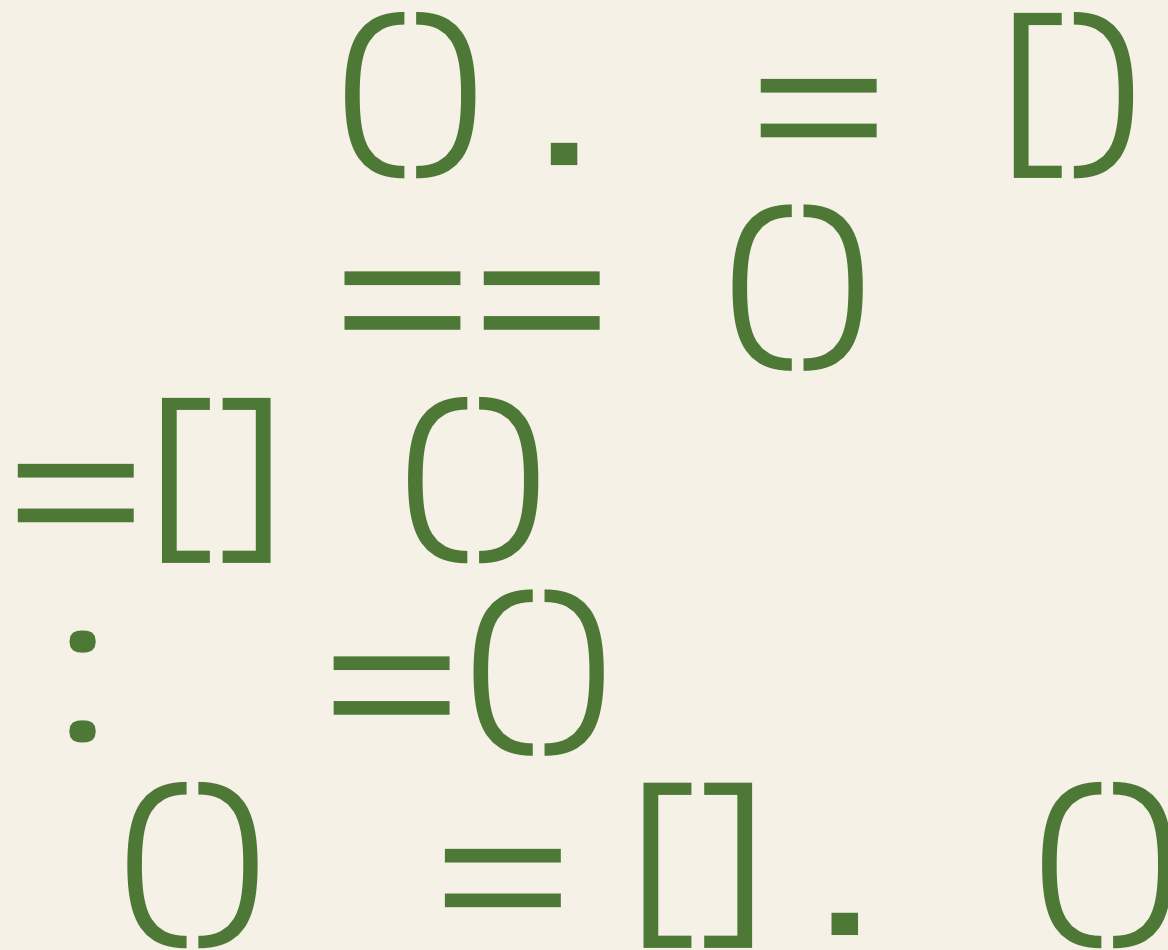
(= \ / =)
(= = () [.])
([:] () ())
= 0 = 0 .
 . [] .

Frokost til 12:45

Opgave 1 – Kortlægning

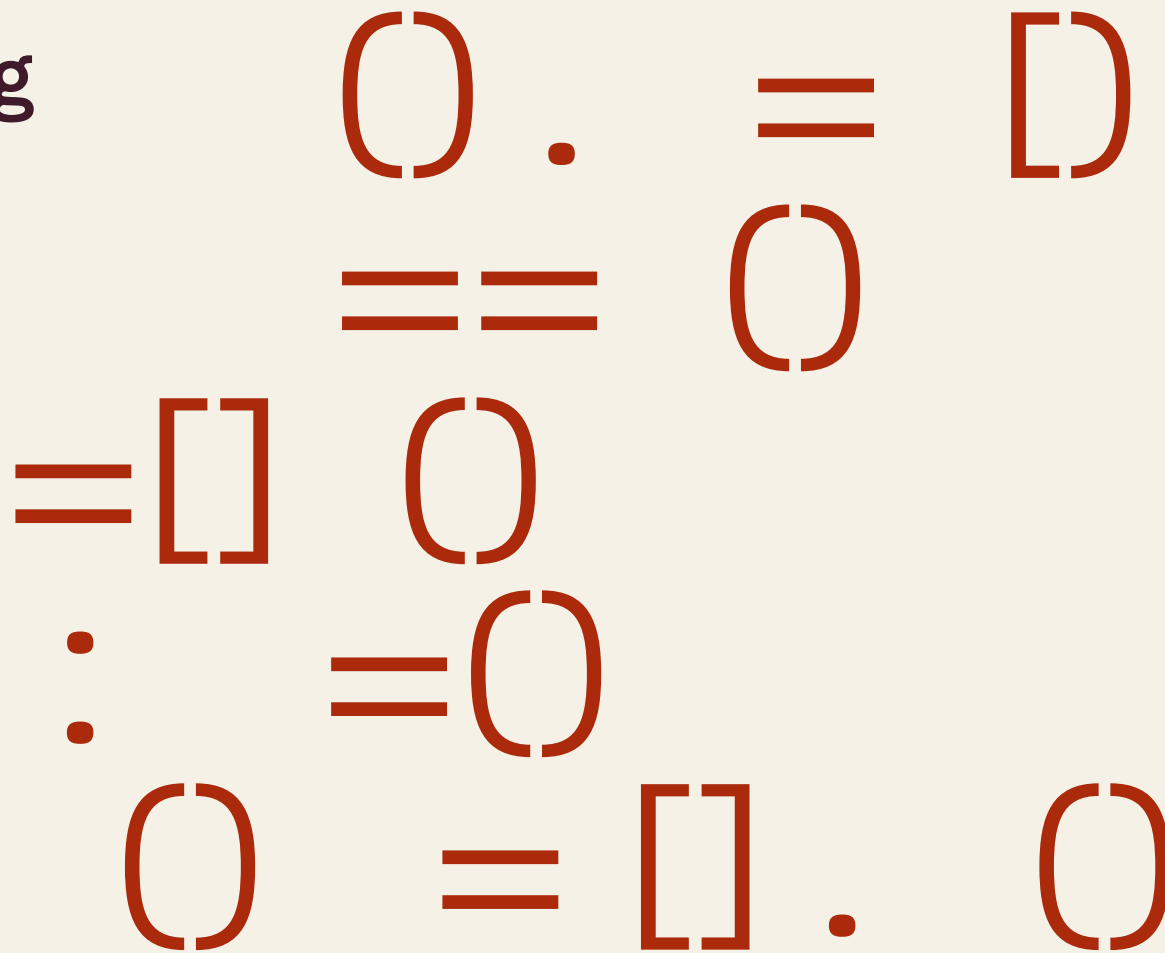
- Formål: At skabe overblik over det vi har og det vi mangler.
- Tre poster:
 1. Det har vi.
 2. Det skal vi have mere af.
 3. Det skal vi ikke have mere af.
- Tænk gerne i:
 - Datasæt
 - Domæner
 - Projekter
 - Evalueringsopgaver/-scenarier

... eller lignende.



Opgave 2 – Backlog og prioritering

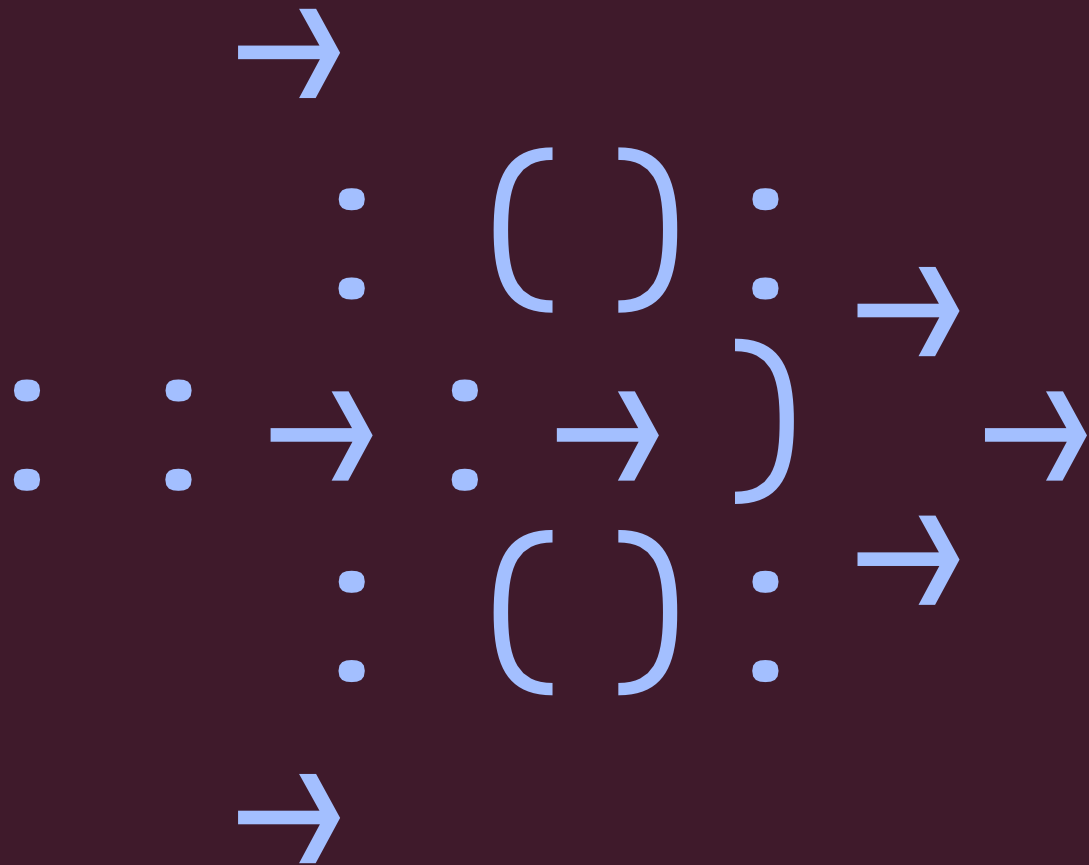
- Formål: at præcisere og prioritere i opgaver, som skal løses.
- Tre kolonner per "opgave":
 1. To do
 - Definer opgaven så specifik som muligt.
 - Tænk gerne her i specifikke datasæt, specifikke evalueringsopgaver/-scenarier, specifikke projekter.
 2. Hvem?
 - Angiv hvilken aktør, der bør løse opgaven.
 - Fx DIGST, men også gerne andre aktører.
 3. Afhængigheder og behov fra andre
 - List forudsætninger for, at aktøren i "Hvem?" kan løse opgaven og giv eventuelt et bud på, hvem der kan hjælpe med dette.
- Undlad i første omgang at prioritere i opgaverne, dette gør vi i plenum efterfølgende.





Afrunding

- Er der nogen åbenlyse næste skridt?
- Er der nogen spørgsmål vi mangler at stille?
- Digitaliseringsstyrelsen udsender en opsamling til deltagerne og på sprogteknologi.dk.
- Mulighed for networking frem til kl. 16:00.



Tak for i dag!