

Tæl din tekst

Introduktion til digital tekstanalyse

> Session 9

> Ulf Dalvad Berthelsen

> udb@cc.au.dk

> <https://github.com/ulfberthelsen>



Dagens program

- Script 1
- Dagens tekster: Aktuelle trends i digital humaniora
- **Gruppearbejde: Pandas 2**
- Forberedelse til session 10

Scripts 1

1. Opvarmning i break out-grupper
2. Diskussion
3. Feedback

Kunstig intelligens (AI) i Digital Humaniora

1. Diskussion i break out-grupper: Hvad er mulighederne og begrænsningerne ved brug af kunstig intelligens i den humanistiske forskning?

Hvad er kunstig intelligens?

- Turing-testen (kommunikativ interaktion): Kan en maskine opføre sig på en måde, så vi ikke kan afsløre at det ikke er et menneske? ... tænk på Philip K. Dick's 'Do Androids Dream of Electric Sheep?' (filmatiseret som Blade Runner)
- “[AI is] the science of making machines do things that would require intelligence if done by men.” (Marvin Minsky)
- Strong vs. weak AI > ‘ægte intelligens’ eller imitation af intelligens?
- Fra “Intelligens” (der minder om menneskelig intelligens) til “Machine learning” (automatiseret mønstergenkendelse og forudsigelse)

Mønstergenkendelse

Kan vi formulere præcise regler, der gør computeren i stand til at identificere alle elementer i række 1 som eksempler på bogstavet 'b' og alle elementer i række 2 som eksempler på bogstavet 'd'?

1: B ***B*** B *B* *B* b **b** *b* *b* **b**

2: ***D*** D *D* *D* D *d* d **d** *d* d

Mønstergenkendelse

Kan vi formulere præcise regler, der gør computeren i stand til at genkende alle elementer i række 1 som eksempler på bogstavet 'B' og alle elementer i række 2 som eksempler på bogstavet 'd'?

1: B *B* B *B* B b **b**

2: *d* D *d* *d* D d d

Superviseret maskinlæring: vi giver computeren feedback og retter fejl. Succeskriteriet er korrekthed.

Ikke-superviseret maskinlæring: computeren finder selv mønstre (som vi ikke nødvendigvis kan se eller forstå). Succeskriteriet er performance.

Kunstig intelligens (AI) i Digital Humaniora

- Hvad kan vi stille op med mønstergenkendelse og forudsigelse i digital humaniora?
- En hel del, fordi (næsten) alle fænomener kan digitaliseres, fx tekst, tale, billeder, music, adfærd osv.
- Vi kan derfor studere alle de klassiske humanistiske områder, dvs. Kunst og kultur i bredeste forstand, ved hjælp af digitale metoder.

Kunstig intelligens (AI) i Digital Humaniora

- Tekst og talegenkendelse > fx til brug for automatiseret digitalisering af ikke digitale materialer (fx identification af enkelt ord i en talestrøm)
- Mikro-mønstergenkendelse i tekster (morfologi og syntaks) > fx til brug for automatiseret POS-tagging, Named Entity Recognition og dependency parsing (fx identifikation af egennavne, subjekter, ledsætninger, firmanavne)
- Makro-mønstergenkendelse i tekster (tekst og kommunikation) > fx til brug for sentiment analyse, genre-genkendelse, kvalitetsurdering, narrativestrukturer, tendenser på tværs af tekster, fx historiske brudflader

Kunstig intelligens (AI) i Digital Humaniora

- Tekst og talegenkendelse > fx til brug for arkivering af digitale materialer (fx identification af enkelte ord)
- Mikro-mønstergenkendelse i tekster (morphologi, automatiseret POS-tagging, Named Entity Recognition, parsing (fx identifikation af egennavne, substantiver, verb))
- Makro-mønstergenkendelse i tekster (tekstklassifikation, sentiment analyse, genre-genkendelse, kvantitative tendenser på tværs af tekster, fx historiske tendenser)

Men det er stadig overladt til os at

- Operationalisere
- Forstå og fortolke
- Finde på nye spørgsmål og anvendelsesmuligheder
- Diskutere etiske problemstillinger
- Vurdere kvalitet og sandhedsværdi
- Vurdere nytteværdien

Case: kønsbias i danske prædikener

1. Diskussion i break out-grupper: Hvilke humanistiske problemstillinger er i optagede af? Diskuter mulighederne for at undersøge disse ved hjælp af digitale metoder

Case: kønsbias i danske prædikener

Eksempel på tendens inden for anvendelsen af digitale metoder:

- Adgang til nye sociale verdner
- Fra fortolkning mod viden – fra 'hvordan skal vi forstå teksterne' til 'hvad kan vi vide om teksterne'
- Fra det (inter)subjektive (smagsdom) til det objektive (sandhed/viden)
- Fokus på og mulighed for at udfordre gamle sandheder og privilegier på et empirisk informeret grundlag
- Paradoksalt i den forstand, at videnskabelige idealer kommer til at understøtte dagsordner, der har rod i miljøer (fx poststrukturalistisk inspirerede), der er stærkt kritiske over for sandhed og kvantitative metoder.

Gruppearbejde

1. Åben Jupyter Notebook
2. Åben filen `Tæl_Pandas_2.ipynb`
3. Bang-teksterne skal gemmes samlet i en mappe, placer mappen i den mappe, hvori I gemmer Notebooken
4. Lav opgaverne i notebook'en i fællesskab, men husk at alle skal afprøve koderne i deres egen notebook!

[illegible]

Forberedelse til session 10 (workshop)

- **Se videoer**
 - Se videoerne om regular expressions
 - Afprøv de eksempler der vises i videoerne
- **Repetition**
 - Repetér Pandas 1 og 2
 - Spørg, hvis der er noget der er uklart!!
- **Forberedelse til eksamensworkshop**
 - Lav en klar problemformulering, enten som spørgsmål eller prosa
 - Lav en detaljeret oversigt i punktform over, hvordan I vil besvare problemformuleringen
 - Hvilke teoretiske begreber vil I bruge
 - Hvad vil I bruge dem til?
 - Uploades som word-dokument på BB senest torsdag.
 - Mundtlig feedback i forbindelse med workshoppen på mandag.
- **Gruppe 3 begynder at forberede et script.**
 - Gruppevejledning efter næste undervisningsgang

[illegible]

> Session 9

```
> udb@cc.au.dk
```