

Dataintensiv Humaniora?

Om Automatisering og Digitaliserings rolle indenfor Humanistiske
Domænekompeticer

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

Center for Humanities Computing
Aarhus University, Denmark

Marts 1, 2019

Outline

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

1 Introduktion

Digitale kompetencer

Introduktion

Digitale kompetencer

2 Humanistisk Informatik

Informatikkens rolle

Et behov for informatik

Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik

Humanioras
transformation

Humanistisk workflow
Datadrevet workflow
Nye udfordringer

3 Humanioras transformation

Humanistisk workflow

Datadrevet workflow

Nye udfordringer

Anvendelseseksempler
Latent semantik
Vedvarende
dynamikker
Adfærdsannotation

4 Anvendelseseksempler

Latent semantik

Vedvarende dynamikker

Adfærdsannotation

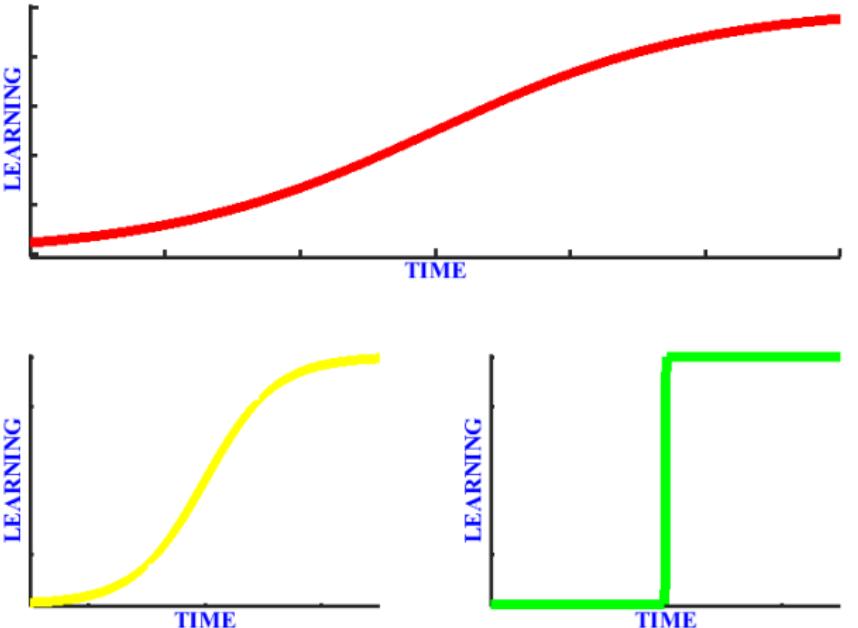


Smutvej til digitale kompetencer

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

- Introduktion
- Digitale kompetencer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Vedvarende dynamikker
- Adfærdsannotation



en KLAASSIK STACK

“(Består af eksplisitte instruktioner til en computer (skrevet af en programmør). Hver linje kode identifierer et punkt i et programrum med en tilsigtet adfærd”

KUNSTIG INTELLIGENS

“Er abstrakt og ligner ikke natursprog, men er vægte i neurale netværk. Mennesker kan ikke programmere direkte i millioner af vægte, i stedet skriver man algoritmer som selv lærer vægtene gennem en optimeringsproces”

Introduktion
Digitale kompetencer

Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik

Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer

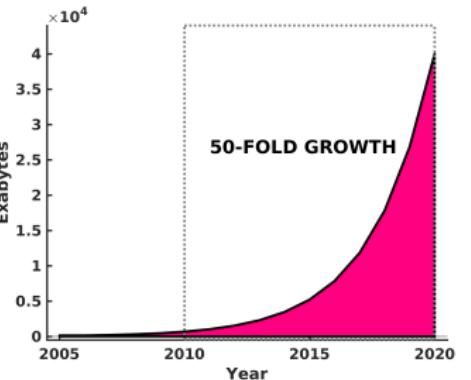
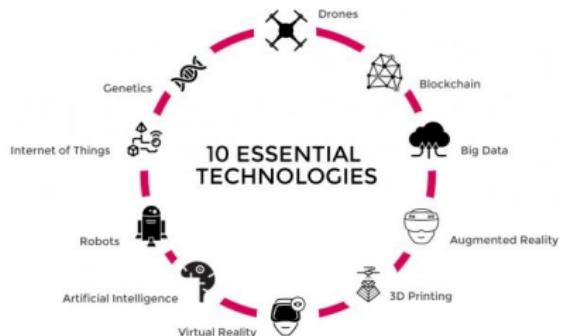
Anvendelseseksempler
Latent semantik
Vedvarende
dynamikker
Adfærdsannotation



Datastrømmen & automatisering

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io



datastrømmen forandrer viden og forståelse i alle domæner

en stor del af datastrømmen er blød og ustruktureret ⇒ for at omsætte

disse data til viden/værdi, skal humanistiske og samfundsvidenskabelige praksisser automatiseres

humanistisk informatik - automatisk informationsbehandling i humaniora,
men skal det ligge under *Humaniora* eller *Informatik*

Introduktion
Digitale kompetencer
Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik
Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Vedvarende
dynamikker
Adfærdsannotation

Humaniora har behov for informatik

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io



Figure 1: Stigning i forskningspublikationer og databaser kræver *computational literacy*. Publikationer relateret til Markusevangeliet > 50K, ~ 16,500 ord i 16 kap. på 11 p.

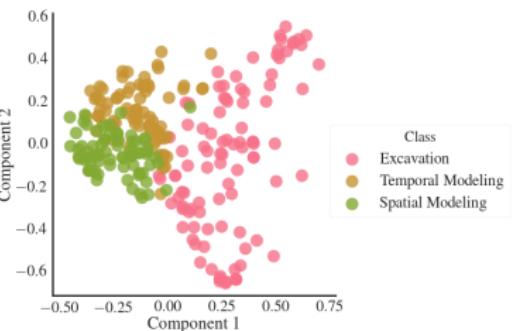


Figure 2: Avanceret humanistisk informatik kan aggregere og projicere store heterogene data til få dimensioner mhp. visuel manipulation

- Introduktion
- Digitale kompetencer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik**
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Vedvarende dynamikker
- Adfærdsannotation

Humanistisk workflow

Dataintensiv
Humaniora?

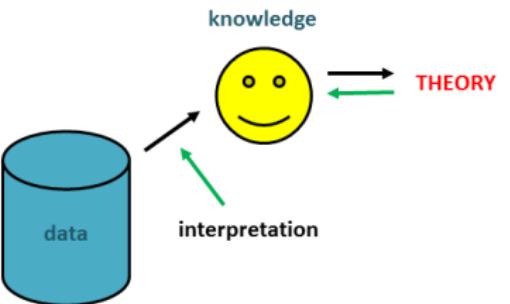
Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

Introduktion
Digitale kompetencer

Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik

Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer

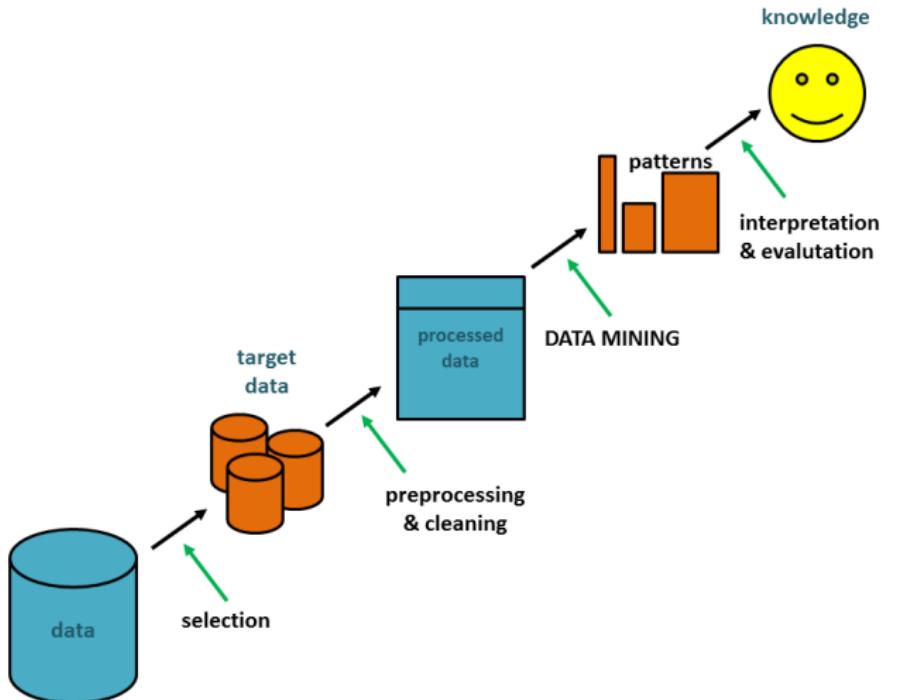
Anvendelseseksempler
Latent semantik
Vedvarende
dynamikker
Adfærdsannotation



Datadrevet workflow

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io



- Introduktion
- Digitale kompetencer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Vedvarende dynamikker
- Adfærdsannotation



Udfordringer for humanistisk informatik

- Bygge, vedligeholde og anvende gode datasæt
- Skabe en kultur, der værdsætter, anvender og prioriterer data
- Anvende software, der understøtter, ikke forhindrer og forsinker
- *Empower* forskere ved at introducere og undersøtte gode processer
- Bruge videnskab og metoder, der skaber adaptive projekter



Litterær optimalitet

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

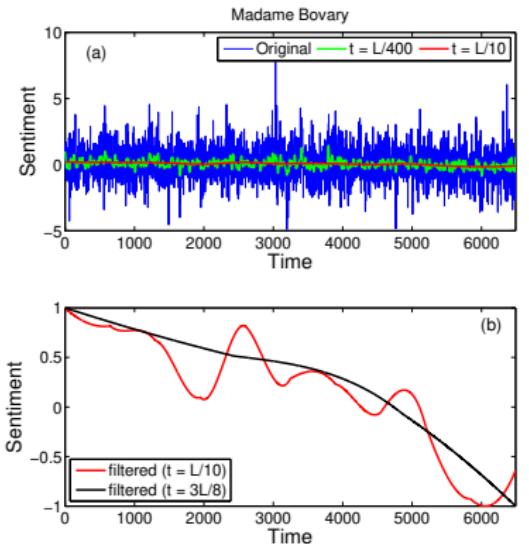
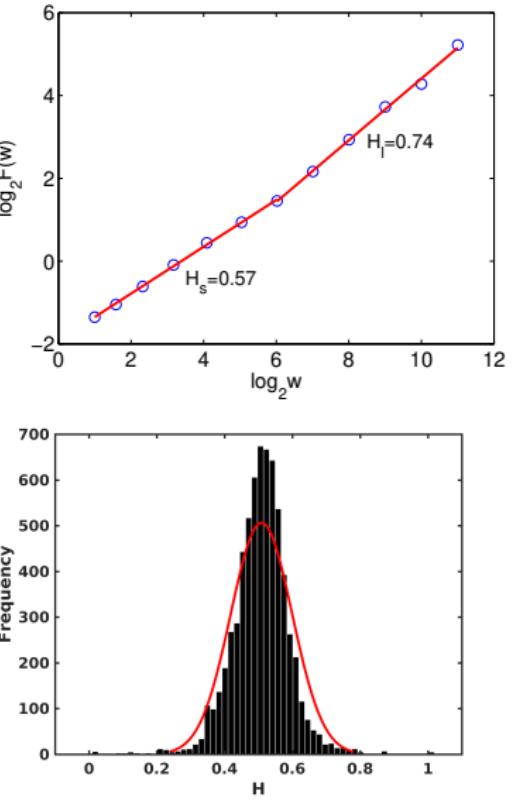


Figure 3: sentimentanalyse & adaptiv filtrering til konstruktion af narrativ vektor, der afspejler læseroplevelsen. Interval for vektorens fraktalstruktur, $0.6 < H \leq 0.8$, som (måske) indikerer litterær optimalitet.



Introduktion
Digitale kompetencer
Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik
Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Vedvarende
dynamikker
Adfærdsannotation

Forandringer i forfatterskaber

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

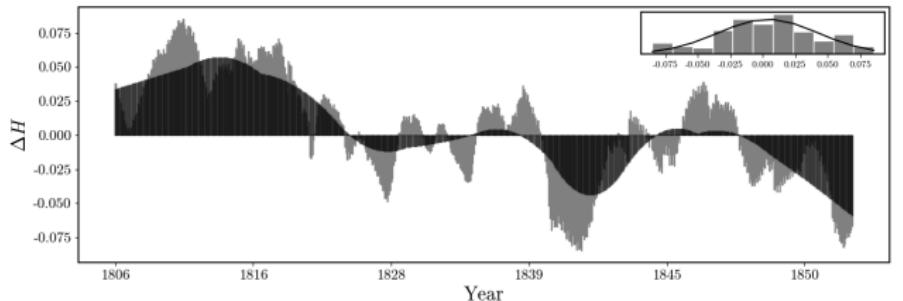


Table 1: Dynamiske faser i N.F.S. Grundtvigs forfatterskab

Time period	Age of onset	$H(X)$	Behavior	Profile
1806-1826	23	$H > 0.5$	<i>persistent</i>	theoretician
1826-1839	43	$H \approx 0.5$	<i>short memory</i>	pragmatic
1839-1845	56	$H < 0.5$	<i>anti-persistent</i>	breakthrough
1845-1848	62	$H \approx 0.5$	<i>short memory</i>	disease
1849-1872	65	$H < 0.5$	<i>anti-persistent</i>	politician

- Introduktion
- Digitale kompetencer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Vedvarende dynamikker
- Adfærdsannotation

Dynamisk forfatterprofilering

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

Introduktion
Digitale kompetencer

Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik

Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer

Anvendelseseksempler
Latent semantik
Vedvarende
dynamikker
Adfærdsannotation

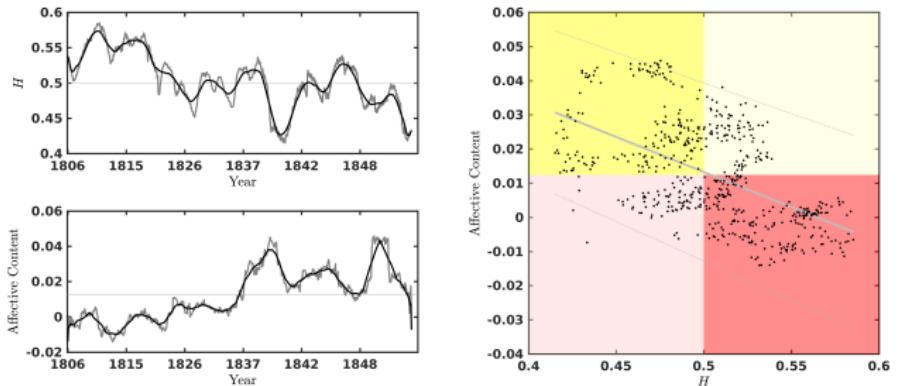


Figure 4: Ved at kombinere vedvarende kreative trends med affektivt indhold, kan vi studere "den forpinte kunstner" i intellektuel historie.



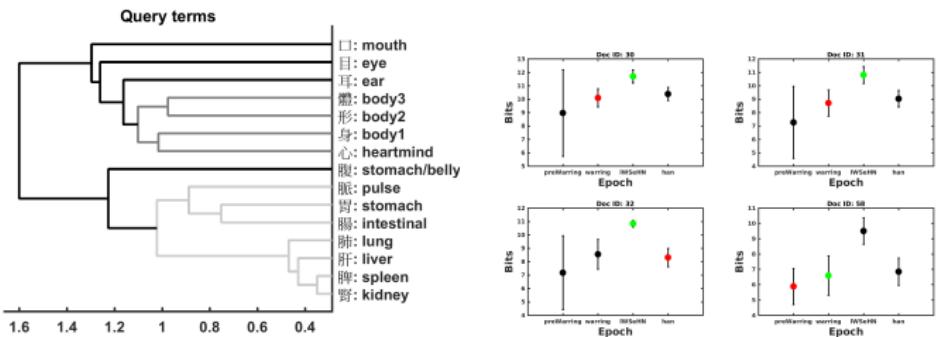


Figure 5: Filosoffer og sinologer debatterer *mind-body* dualisme i klassisk kinesisk filosofi. I samarbejde med domæneeksperter, identificerede vi et hierarkisk dualistisk rum baseret på en latent semantisk model. En model blev senere brugt til identifikation af teksters oprindelse.

Introduktion
Digitale kompetencer
Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik
Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Vedvarende
dynamikker
Adfærdsannotation

Slingerland, E., Nichols, R., Nielbo, K., & Logan, C. (2017). The Distant Reading of Religious Texts: A Big Data Approach to Mind-Body Concepts in Early China. *Journal of the American Academy of Religion*, 85(4), 985–1016.

Nichols, R., Slingerland, E., Nielbo, K., Bergeton, U., Logan, C., & Kleinman, S. (2018). Modeling the Contested Relationship between Analects, Mencius, and Xunzi: Preliminary Evidence from a Machine-Learning Approach. *The Journal of Asian Studies*, 77(01), 19–57.

Overgange i middelalderhistorie

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

- historikere debatterer overgange i tekstsamlinger
- Saxos *Gesta Danorum* (ca 1200) semifiktionel danmarkshistorie
- overgang mellem bog 8 or 9?
- punkt eller gradvis?
- traditionelle ordbaserede modeller giver ikke klart svar
- latent semantisk model trænet på sætninger
- detektion af forandringer og change detection og gentagelser kan identificere en overgang i bog 9

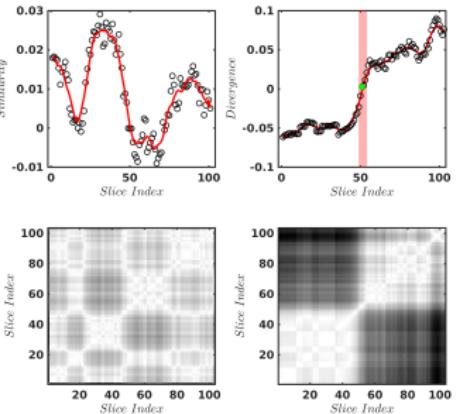


Figure 6: Cosinusafstand and relativ entropy for dokument-ord vektor og guided LDA model.

Introduktion
Digitale kompetencer
Humanistisk Informatik
Informatikkens rolle
Et behov for informatik
Humanioras transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Vedvarende dynamikker
Adfærdsannotation

%indenrigsnyheder er indeks for populisme

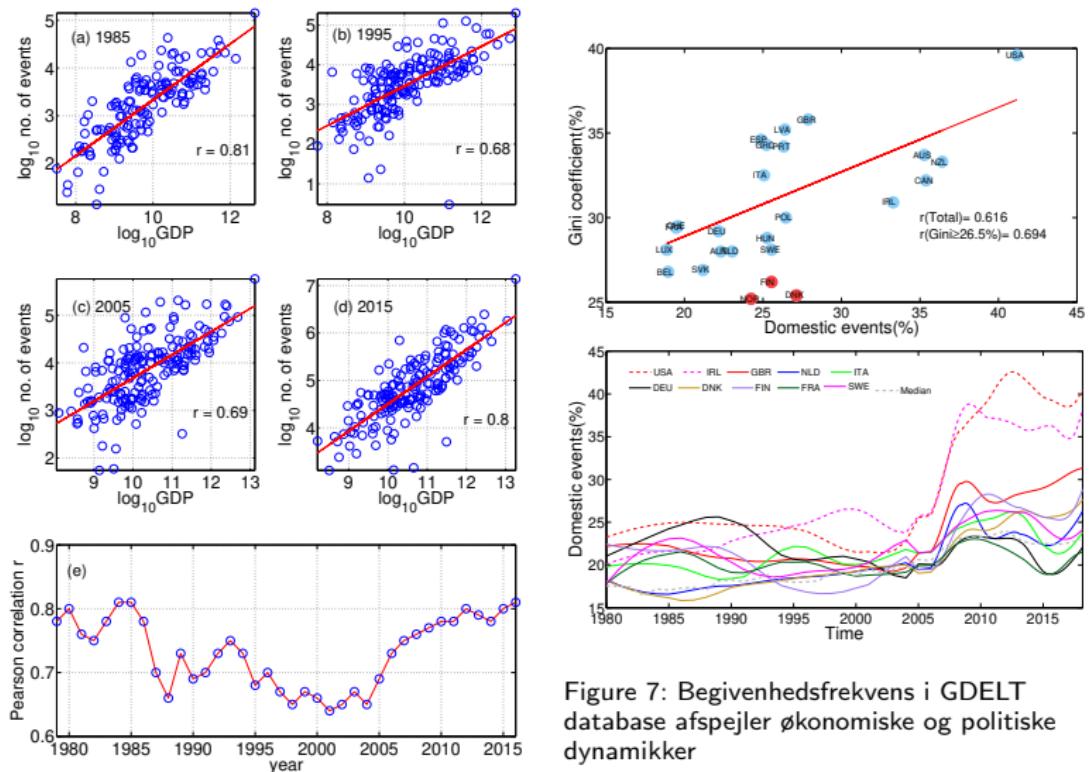


Figure 7: Begivenhedsfrekvens i GDELT database afspejler økonomiske og politiske dynamikker

Forbrug og kulturel hukommelse

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

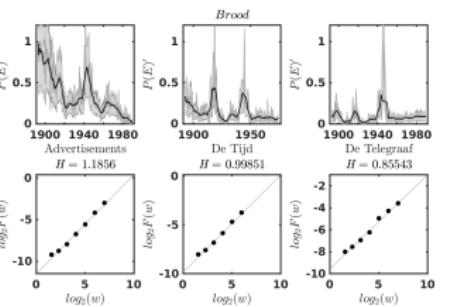
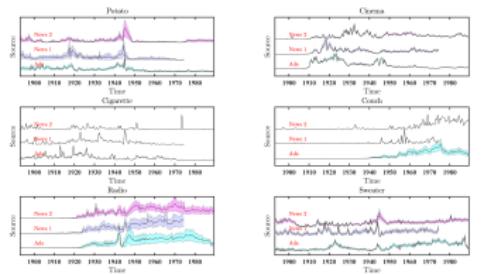


Figure 8: Historikere og medieforskere teoretiserer om kausale relationer mellem offentlig diskurs og reklamer. Kausal modellering af nøgleord indikerer at for nogle kategorier "former reklamer samfundet", mens andre kategorier blot "afspejler behov". Reklamer udviser mere kaotisk adfærd en offentlig diskurs \sim en proxy for kulturel hukommelse.

- Introduktion
- Digitale kompetencer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Vedvarende dynamikker
- Adfærdsannotation

Decision support til OCD

Dataintensiv
Humaniora?

Kristoffer L Nielbo
kln@cas.dk
knielbo.github.io

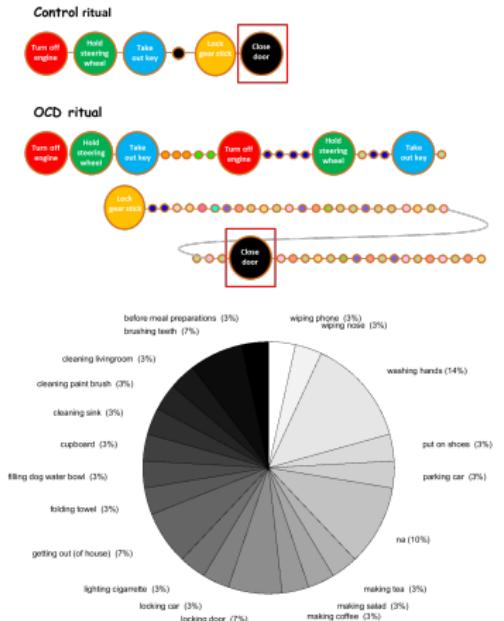


Figure 9: Handlingsdatabase annoteret med XT Observer for OCD, komorbiditet og kontrol.¹

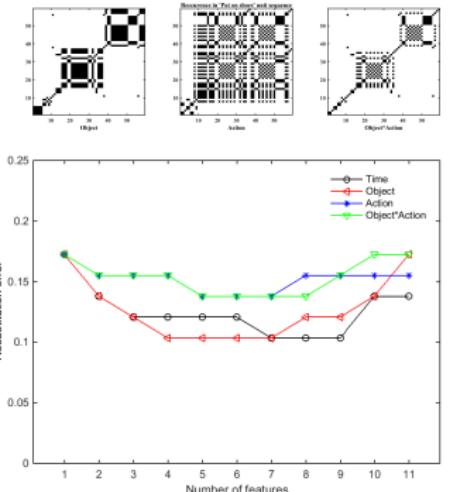


Figure 10: Binomial klassifikation (OCD vs. kontrol) på testdata.

¹Zor, R., Hermesh, H., Szechtman, H., & Eilam, D. (2009). Turning order into chaos through repetition and addition of elementary acts in obsessive-compulsive disorder (OCD). World Journal of Biological Psychiatry, 10(4.2), 480–487.

Introduktion
Digitale kompetencer
Humanistisk Informatik
Informatikkens rolle
Et behov for informatik
Humanioras transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Vedvarende dynamikker
Adfærdsannotation

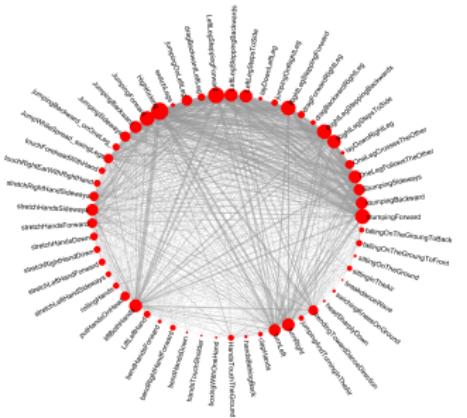


Figure 11: Handlingsdatabase annoteret med XT Observer for rituel dans under **zulubryllup** for to generationer med kvindelige og mandelige deltagere.

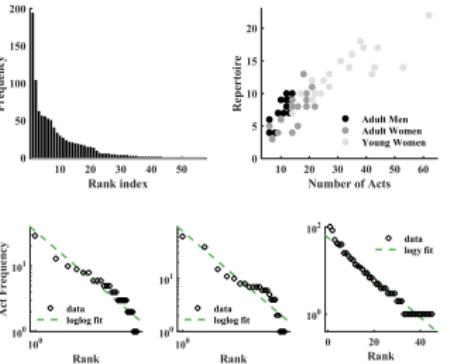


Figure 12: Bevægelsesleksikon følger Heaps and Zipfs love kendt fra andre **kommunikative systemer**.

Introduktion
Digitale kompetencer
Humanistisk Informatik
Informatikkens rolle
Et behov for informatik
Humanioras transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Vedvarende dynamikker
Adfærdsannotation

TAK

kln@au.dk
knielbo.github.io

slides: http://knielbo.github.io/files/kln_fip19.pdf

& tak til

Mads Rosendahl Thomsen, Comparative Literature, School of Communication and Culture, Aarhus University, DK

Jianbo Gao and Bin Liu, Institute of Complexity Science and Big Data, Guangxi University, CHN

Culture Analytics @ Institute of Pure and Applied Mathematics, UCLA, US

