

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
- Software paradigmer

Introduktion
Digitale kompetencer
Software paradigmer

2 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
Persistent dynamics
Behavioral annotation

4 Anvendelseseksempler

- Latent semantik
- Persistent dynamics
- Behavioral annotation

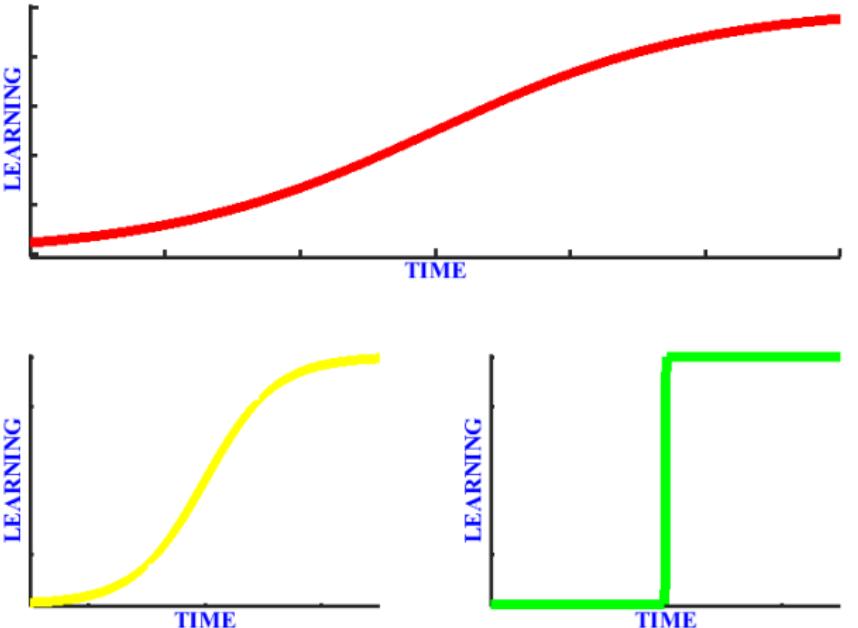


Smutvej til digitale kompetencer

Dataintensiv
Humaniora?

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

- Introduktion
- Digitale kompetencer
- Software paradigmær
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Persistent dynamics
- Behavioral annotation



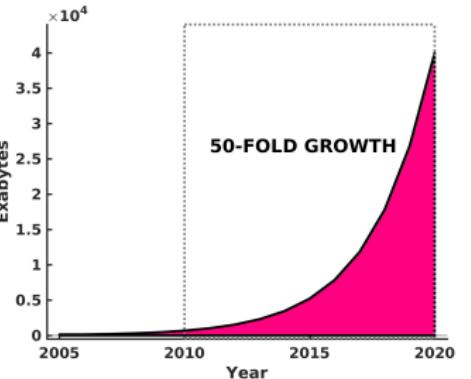
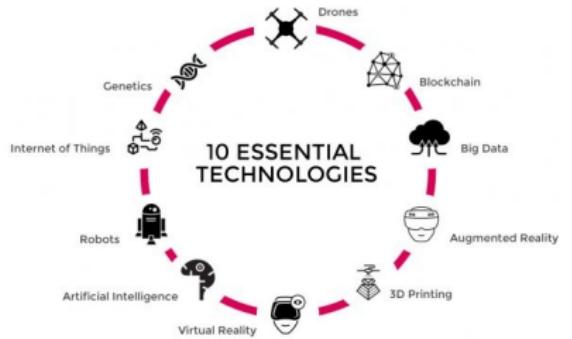
- "(en klassisk *stack*) består af eksplisitte instruktioner til en computer (skrevet af en programmør). Med hver linje kode identifierer programmøren et punkt i et programrum med en tilsigtet adfærd"
 - "det nye paradigme er mere abstract og ligner ikke natursprog, men er vægte i neurale netværk. Mennesker kan ikke programmere direkte i millioner af vægte, i stedet skriver man algoritmer som selv lærer vægtene gennem en optimeringsproces"
- ⇒ *coding-for-all* initiativer er underspecificerede på dette område
⇒ nødvendigt at undervise i håndtering og analyse af data



Datastrømmen & automatisering

Dataintensiv
Humaniora?

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



datastrømmen forandrer videnstilegnelse og forståelse i samtlige domæner

en stor del af disse data er bløde og ustukturerede ⇒ for at omsætte

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

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

Introduktion
Digitale kompetencer
Software paradigmer
Humanistisk Informatik
Informatikkens rolle
Et behov for informatik
Humanioras transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Persistent dynamics
Behavioral annotation

Humaniora har behov for informatik

Dataintensiv
Humaniora?

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



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

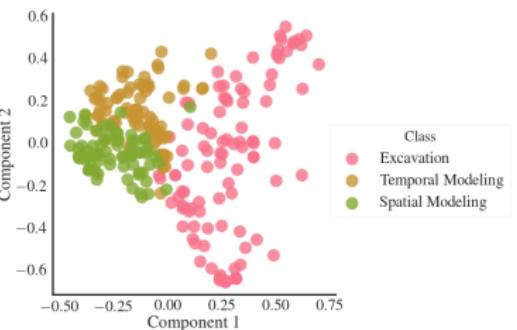


Figure 2: Avanceret humanistisk informatik kan aggregere og projicere store heterogene data til få dimensioner mhp. visuel manipulation

- Introduktion
- Digitale kompetencer
- Software paradigmer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anwendungseksempler
- Latent semantik
- Persistent dynamics
- Behavioral annotation

Humanistisk workflow

Dataintensiv
Humaniora?

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

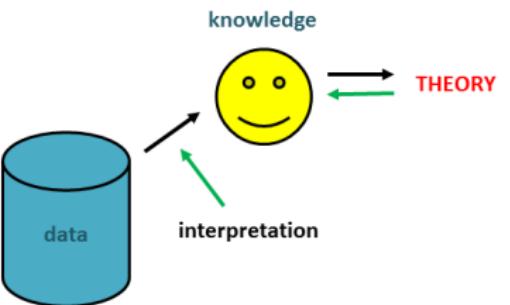
Introduktion
Digitale kompetencer
Software paradigmer

Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik

Humanioras
transformation

Humanistisk workflow
Datadrevet workflow
Nye udfordringer

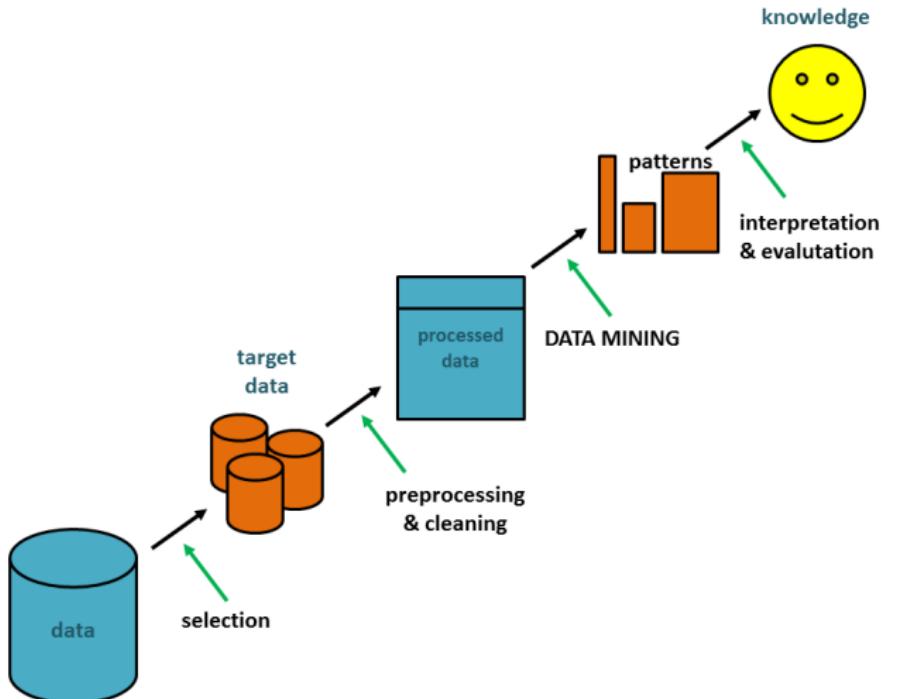
Anvendelseseksempler
Latent semantik
Persistent dynamics
Behavioral annotation



Datadrevet workflow

Dataintensiv
Humaniora?

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



- Introduktion
- Digitale kompetencer
- Software paradigmer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow**
- Nye udfordringer
- Anvendelsesksempler
- Latent semantik
- Persistent dynamics
- Behavioral annotation



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 og introducere og undersøtte gode processer
- bruge videnskab og metoder, der skaber adaptive projekter

⇒ **data, kultur, mennesker & processer, software, og videnskab & metoder**



Literær optimalitet

Dataintensiv
Humaniora?

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

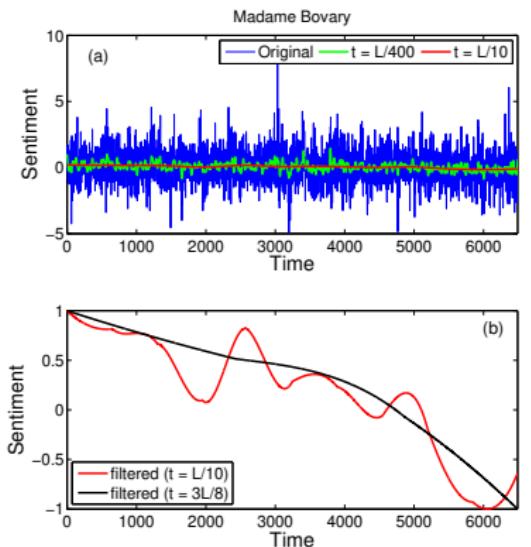
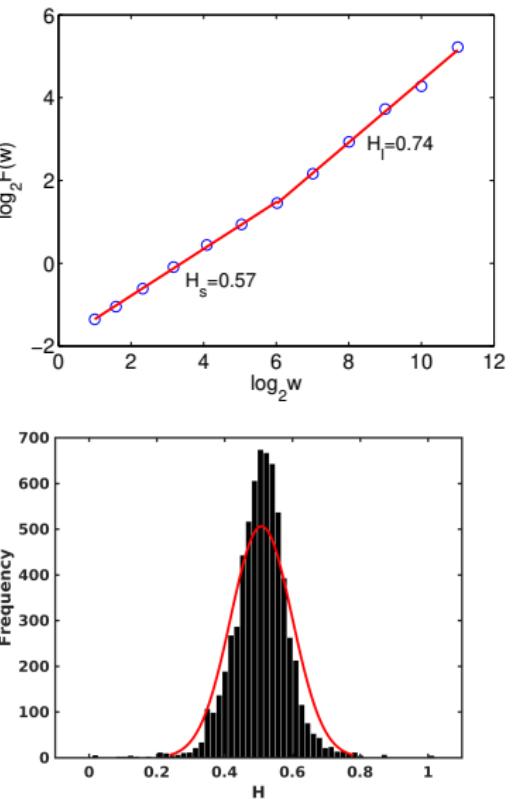


Figure 3: sentimentanalyse & adaptivt
filtrering til konstruktion af narrativ vektor,
der afspejler læseroplevelsen. Interval for
vektorens fraktalstruktur, $0.6 < H \leq 0.8$,
der indikerer literær optimalitet.



- Introduktion
- Digitale kompetencer
- Software paradigmer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Persistent dynamics
- Behavioral annotation

Author change points

Dataintensiv
Humaniora?

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

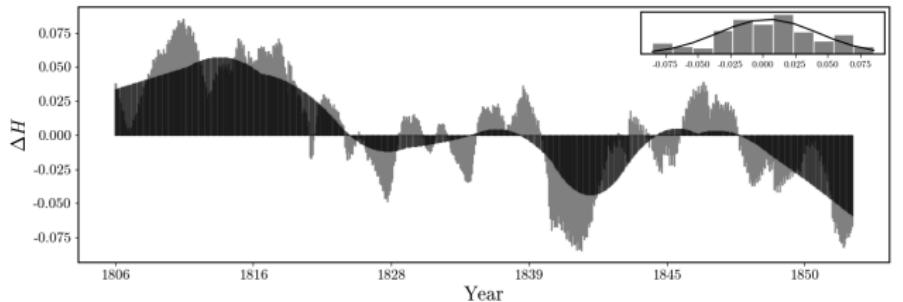


Table 1: Centrale (dynamiske) faser i N.F.S. Grundtvigs forfatterskab

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
Software paradigmer
Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik
Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Persistent dynamics
Behavioral annotation

Dynamic author profiling

Dataintensiv
Humaniora?

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

Introduktion
Digitale kompetencer
Software paradigmer

Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik

Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer

Anvendelseseksempler
Latent semantik
Persistent dynamics
Behavioral annotation

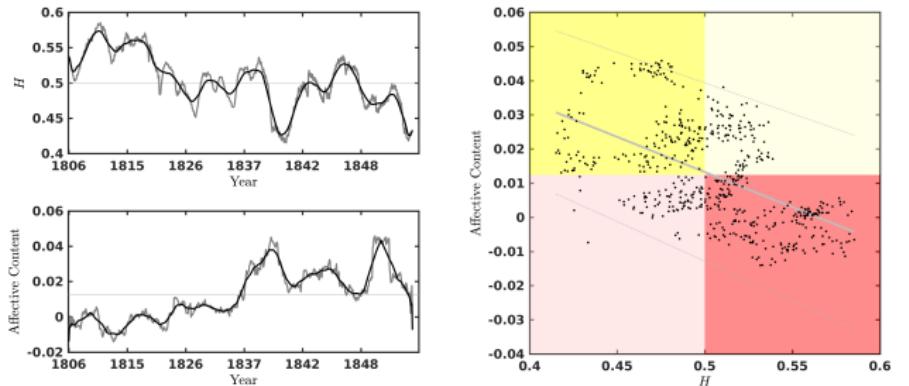


Figure 4: Combining persistent entropic trends with sentiment analysis and causal modeling, we can study “**the tormented artist**” phenomena in intellectual history.



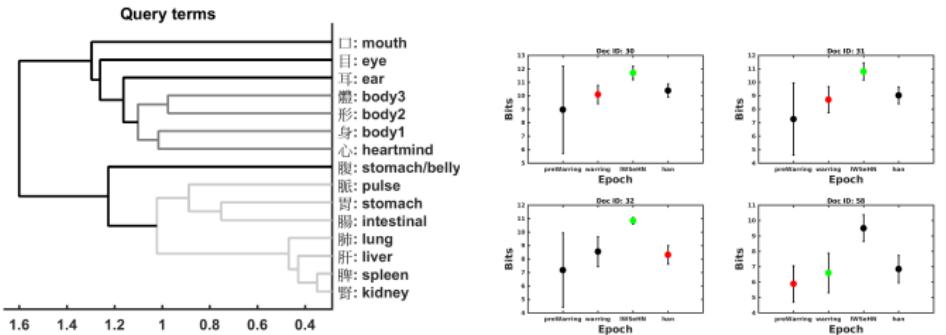


Figure 5: Filosoffer og sinologer debatterer *mind-body dualism* i klassisk filosofi
 Philosophers and sinologists have been debating the existence of **mind-body dualism** in classical Chinese philosophy. With domain experts, we identified a **hierarchical dualistic space** based on latent semantic models. One model (LDA) was further utilized to predict class of origin for **controversial texts** slices.

Introduktion
 Digitale kompetencer
 Software paradigmer
 Humanistisk Informatik
 Informatikkens rolle
 Et behov for informatik
 Humanioras transformation
 Humanistisk workflow
 Datadrevet workflow
 Nye udfordringer
 Anvendelseseksempler
Latent semantik
 Persistent dynamics
 Behavioral annotation

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.

Medieval history and novelty detection

Dataintensiv
Humaniora?

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

- historians debate historical transitions
- Saxo's *Gesta Danorum* c. 1200 CE history of the Danish royal dynasty
 - transition between book 8 or 9?
 - transition point or gradual?
 - traditional word-level representation is ambivalent
 - latent semantic model was trained over sentence windows
 - change detection and recurrence plot used to identify phase transition centered in book 9

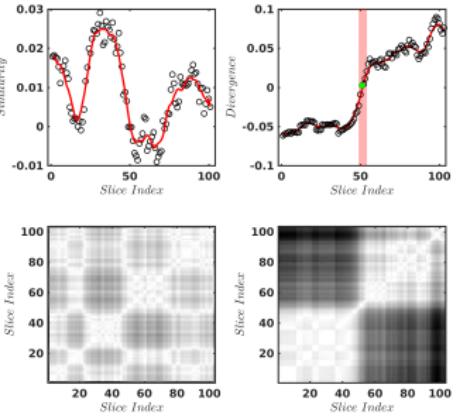


Figure 6: Cosine distance and KLD for TD high-rank vector space and guided LDA model respectively.

Introduktion
Digitale kompetencer
Software paradigmer
Humanistisk Informatik
Informatikkens rolle
Et behov for informatik
Humanioras transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Persistent dynamics
Behavioral annotation

Introduktion
Digitale kompetencer
Software paradigmer
Humanistisk Informatik
Informatikkens rolle
Et behov for
informatik
Humanioras
transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Persistent dynamics
Behavioral annotation

%domestic events indicates populism

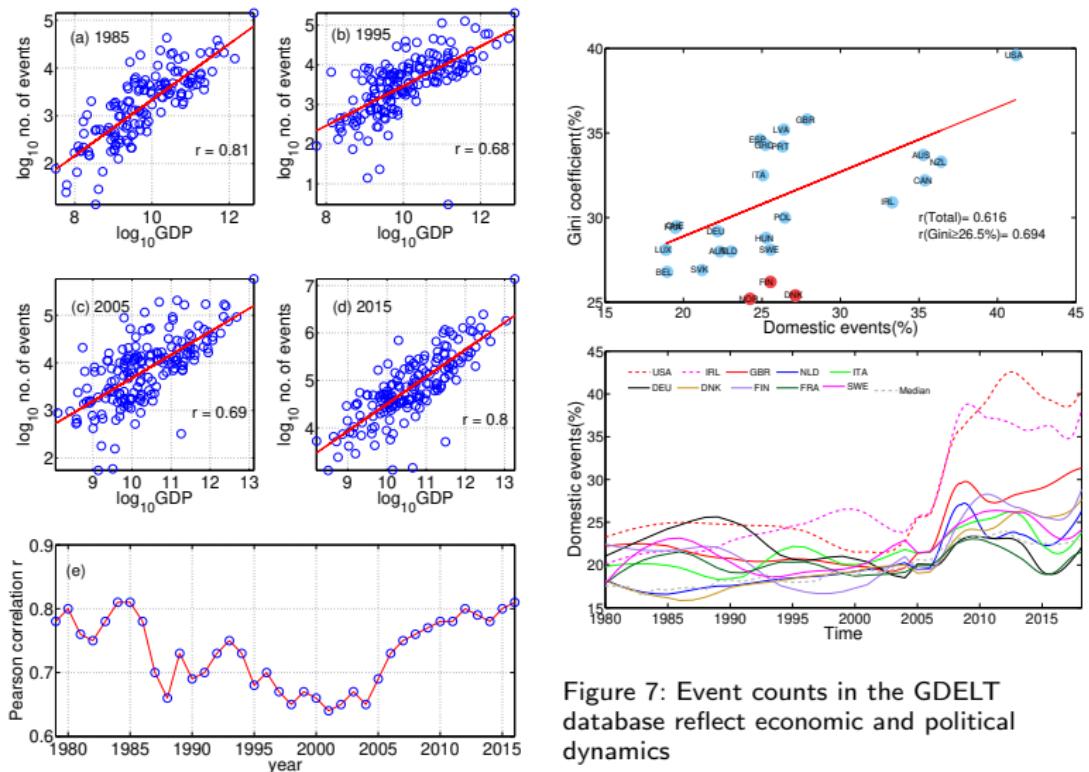


Figure 7: Event counts in the GDELT database reflect economic and political dynamics

Consumer goods and cultural memory

Dataintensiv
Humaniora?

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

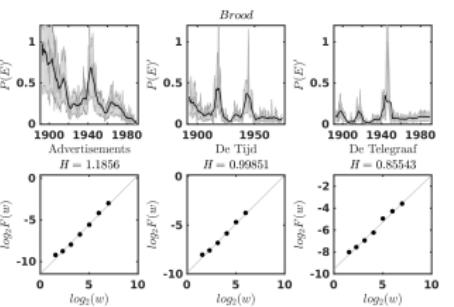
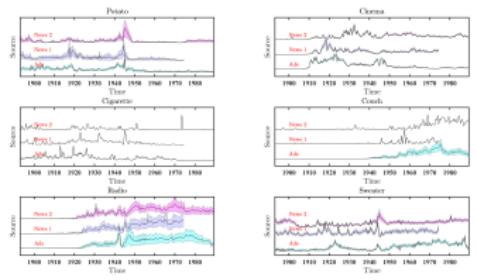


Figure 8: Historians and media researchers theorize about the causal dependencies between public discourse and advertisement. Causal modeling of keyword frequencies (from seedlists) indicated that for some categories 'ads shape society', while other categories merely 'reflect'. Advertisements show a faster decay (on-off intermittent behavior) than public discourse (long-range dependencies) - a proxy for cultural memory.

- Introduktion
- Digitale kompetencer
- Software paradigmer
- Humanistisk Informatik
- Informatikkens rolle
- Et behov for informatik
- Humanioras transformation
- Humanistisk workflow
- Datadrevet workflow
- Nye udfordringer
- Anvendelseseksempler
- Latent semantik
- Persistent dynamics
- Behavioral annotation

Decision support for OCD

Dataintensiv
Humaniora?

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

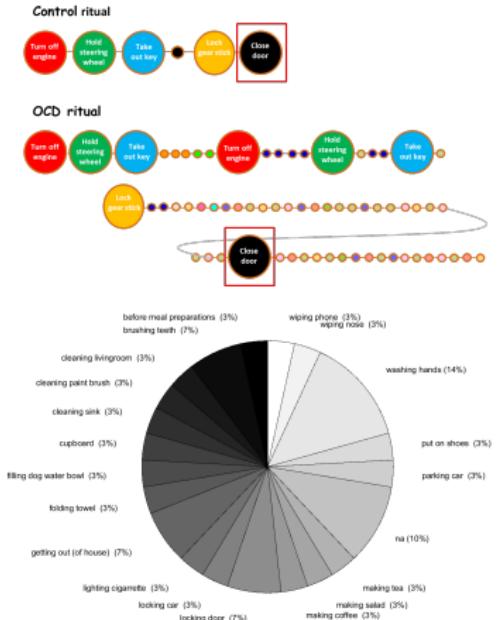


Figure 9: Event logging database annotated with Observer XT for OCD, comorbid, and control.¹

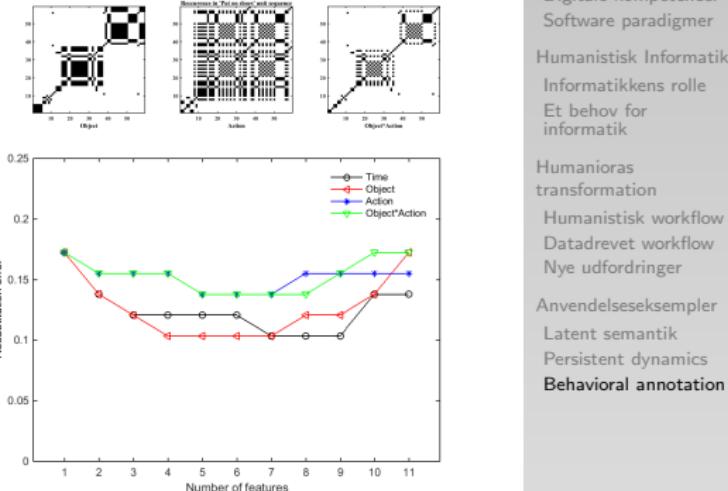


Figure 10: Binomial classifier (OCD vs. control) on unseen data.

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
Software paradigmer
Humanistisk Informatik
Informatikkens rolle
Et behov for informatik
Humanioras transformation
Humanistisk workflow
Datadrevet workflow
Nye udfordringer
Anvendelseseksempler
Latent semantik
Persistent dynamics
Behavioral annotation

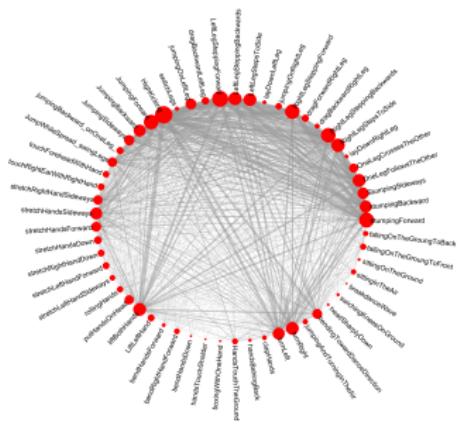


Figure 11: Event logging database annotated with Observer XT for ritualistic dance during **Zulu weddings** in two generations with male/female participants.

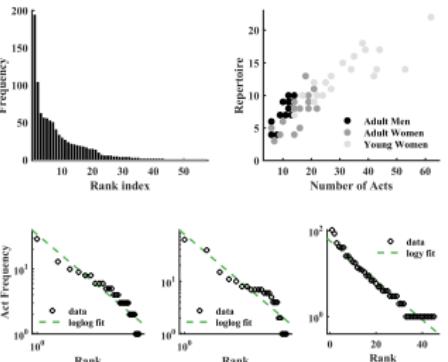


Figure 12: Movement inventory follow Heaps' and Zipf's laws known from other **communicative systems**.

TAK

knielbo@sdu.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

