

# Digital Humanities & Humanities Data

Kristoffer L Nielbo

kln@cas.au.dk

DAI|IMC|AARHUS UNIVERSITY

**does it make sense to talk about a unified DH anymore?**

## DIGITAL HUMANITIES|DH

---

- the *conjunction* of the digital and the humanities
- *redrawing boundary* lines traditional academic boundaries
- expanding the audience and *social impact of scholarship* in the humanities
- developing *new forms of inquiry* and knowledge production
- training *future generations* of humanists through hands-on, project-based learning
- increase *visibility* of humanistic inquiry

*DH has gone through several distinct transformations during the last 3 decades*

---

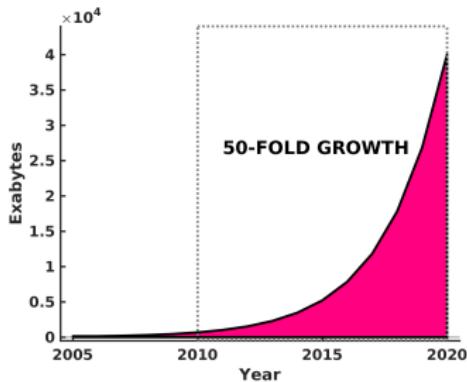
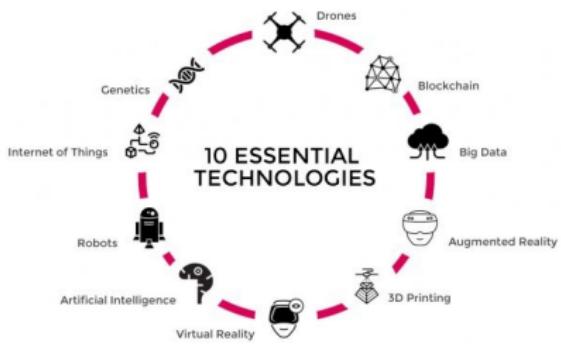
**HUMANITIES COMPUTING** – computational tools in traditional humanities

**DIGITAL HUMANTIES 2.0** – digital culture, activism and critique

**HUMANITIES DATA** – data-intensive research in SSH domains

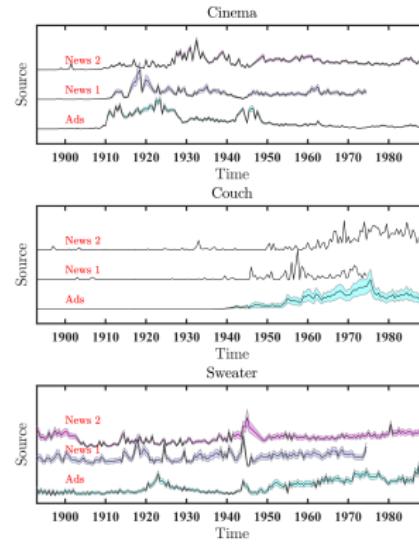
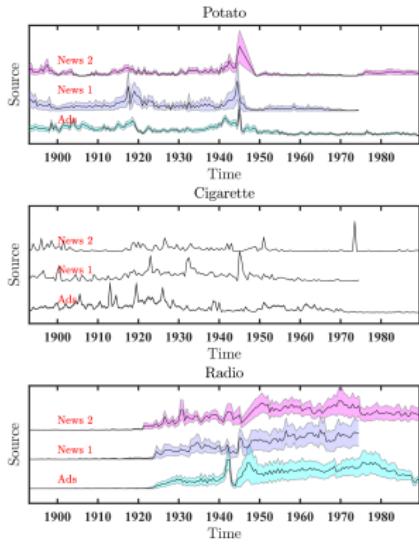
---





*'Digitization and digital media have generated a rapid proliferation of data that is unprecedented in the history of man. This **digital surge** is transforming knowledge discovery and understanding in every domain of human inquiry. Digital research, computing, data management, and data-intensive methods will therefore become integral parts of internationally-leading research in the humanities and arts.'*

(Digital Arts Strategy 2016)



**humanities data example – humanistic domain knowledge (history og medie studies), data engineering, programming, statistics & visualisation**

## what is so new about DATA (or we always had data)?

Data



Information



Presentation



Knowledge



Data



Information



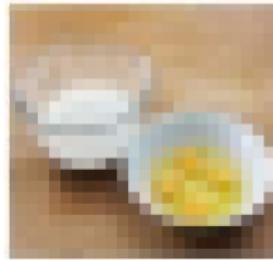
Presentation



Knowledge



Data



Information



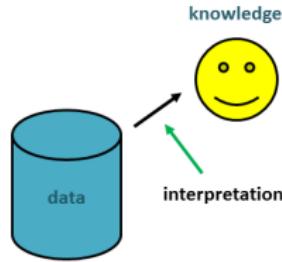
Presentation

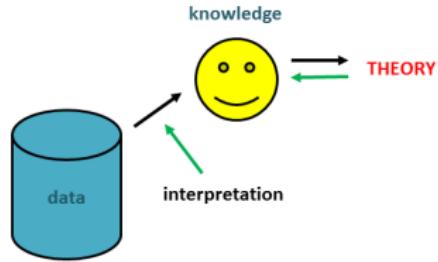


Knowledge



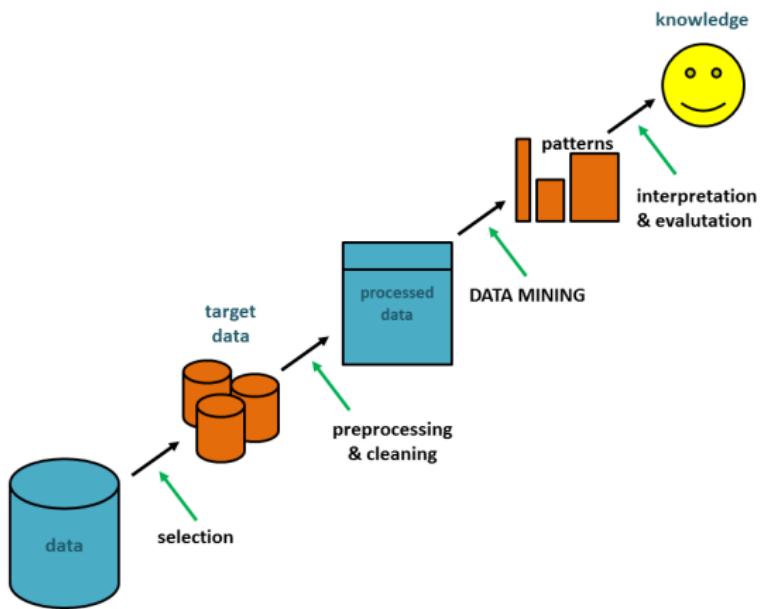
## **DATA modifies our workflow?**

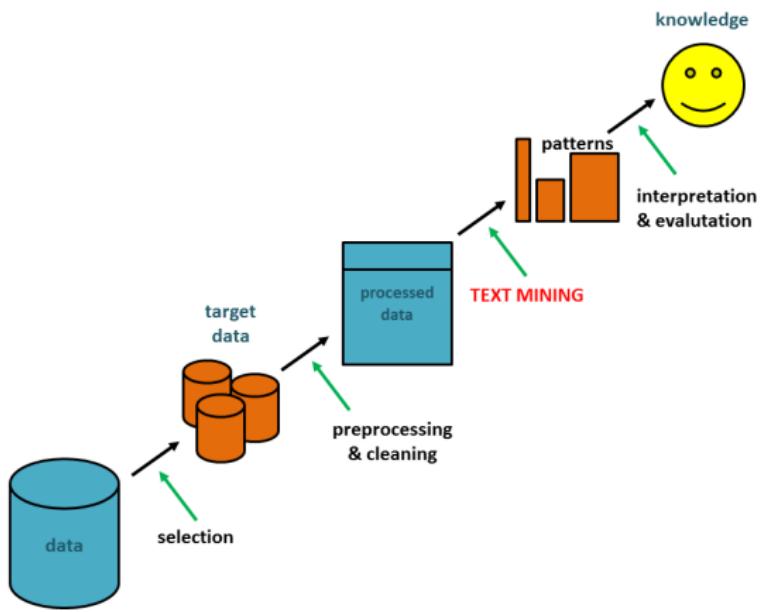




knowledge







## **examples of data-intensive methods in DH**

## PROBLEM #1

---

a document is an ordered set of words that (at least in part) expresses the **cognitive and affective states** of the author

we want a **automatized method** that transfers psychological scales to documents and maintain validity and reliability

preferably, the method should be **scalable** both in terms of quantity and context

a **dictionary** can identify keywords in a collection of documents and apply a sentiment function

```
1 'Did Crooked Hillary help disgusting (check out sex tape and past) Alicia M become a U.S. citizen  
2 so she could use her in the debate?'  
3  
4 Positive sex, citizen  
5 Negative crooked, hillary, disgusting, out  
6 Sentiment Score (2+1) + (-2-1-3-1) = -4  
7 Sentiment Polarity Negative  
8 Overall Score Sum of all sentence scores
```

a sentiment vector is simply a vector of keyword frequencies weighted by sentiment scores

**sentiment analysis** a set of methods for extracting the (primarily) affective components from unstructured data

utilize existing dictionaries to avoid tedious manual coding and validation\*

three general approaches:

- **dictionary-based methods** (word counting)
- supervised learning (ML)
- unsupervised learning (ML)

## PROBLEM #2

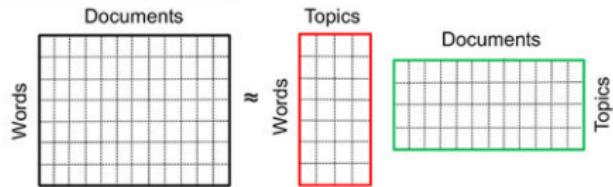
a document is a structured\* or non-random collection of words, but who or what is the structuring agent?

to **avoid manual modeling**, we want method that we can throw ++documents at and then it will sort things out

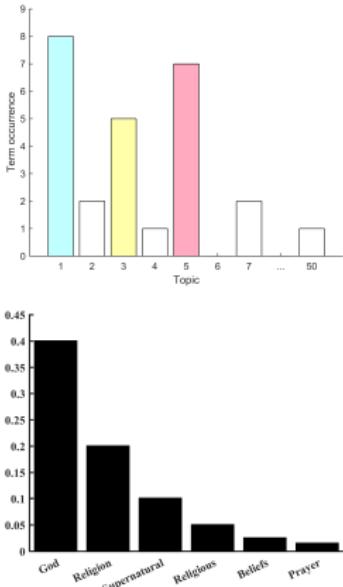
preferably, the output should exhibit some degree of similarity with human text comprehension

we can extract latent topics that generated the documents by reverse engineering the process *words & doc  $\Rightarrow$  topics*

decompose an n-by-d word document matrix into two matrices: a n-by-k **word topic matrix** and an k-by-d **topic document matrix**



to “generate” a document, choose a distribution over topics, sample a topic ( $k$ -by- $d$  matrix), then a word from this topic ( $n$ -by- $k$  matrix) and repeat



**ABSTRACT**—We present two studies aimed at resolving experimentally whether **religion** increases prosocial behavior in the anonymous dictator game. Subjects allocated more money to anonymous strangers when **God** concepts were implicitly activated than when neutral or no concepts were activated. This effect was at least as large as that obtained when concepts associated with secular moral institutions were primed. A trait measure of self-reported religiosity did not seem to be associated with prosocial behavior. We discuss different possible mechanisms that may underlie this effect, focusing on the hypotheses that the religious prime had an **ideomotor effect** on generosity or that it activated a *felt presence of supernatural watchers*. We then discuss implications for theories positing religion as a facilitator of the emergence of early large-scale societies of **cooperators**.

Many theorists have suggested that the cognitive availability of omniscient and omnipresent **supernatural agents** has had a dramatic impact on the development of large-scale human societies. The imagined presence of such agents, along with emotional ritual and costly commitment to the social group they govern, may have been the major development that allowed genetically unrelated individuals to interact in cooperative ways (e.g., Atman & Norenzayan, 2004; Irons, 1991; Sosis & Ruffle, 2004). The research reported in this article experimentally investigated this link between two broad classes of culturally widespread phenomena of interest to social science—**religious beliefs** and **cooperative behavior** among unrelated strangers.

Although anecdotes documenting religion's prosocial and antisocial effects abound, the empirical literature has produced mixed results regarding religion's role in prosocial behavior.

Sosis and Ruffle (2004) examined levels of generosity in an **experimental** cooperative pool game in religious and secular kibbutzim in Israel and found higher levels of cooperation in the religious ones, and the highest levels among religious men who engaged in daily communal **prayer**. Batson and his colleagues (Batson et al., 1999; Batson, Schoenrade, & Ventis, 1993) have shown that although religious people report more explicit willingness to care for others than do nonreligious people, controlled laboratory measures of **altruistic behavior** often fail to corroborate this difference. Furthermore, when studies demonstrate that helpfulness is higher among more devoted people, this finding is typically better explained by egoistic motives such as seeking praise or avoiding guilt, rather than by higher levels of compassion or by a stronger motivation to benefit other people.

However insightful these findings are, research on religion and prosocial behavior has been limited by its overwhelming reliance on **correlational designs**. If religiosity and prosocial behavior are found to be correlated, it is just as likely that having a prosocial disposition causes one to be religious, or that some third variable such as gull proneness or dispositional empathy causes both cooperative behavior and religiosity, as that religious beliefs somehow cause prosocial behavior. Only rarely have studies induced **supernatural beliefs** to examine them as a causal factor. Bering (2003, 2006) inhibited 3-year-old children's tendencies to cheat (*i.e.*, open a "forbidden box") by telling them that an invisible agent ("Princess Alice") was in the room with them. In a different study, college students who were casually told that the ghost of a dead graduate student had been spotted in their private testing room were less willing to cheat on a computerized spatial-reasoning task than were those told nothing (Bering, McLeod, & Shackelford, 2005). These studies suggest that explicit thoughts of **supernatural agents** curb cheating behavior.

In the research reported here, we examined the effect of **God** concepts specifically on selfish and prosocial behavior. Our research design was novel in two ways. First, we introduced an

## PROBLEM #3

word meaning is **dependent on word context** and words in similar contexts have similar meanings

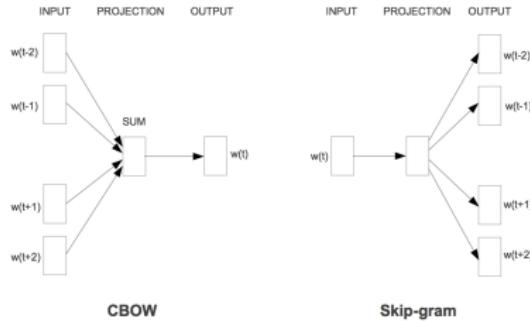
to increase computational efficiency and obtain distributed representations for words, we want a methods that can learn every association from a large set of documents without human interference

preferably, the method should **emulate human concept learning**

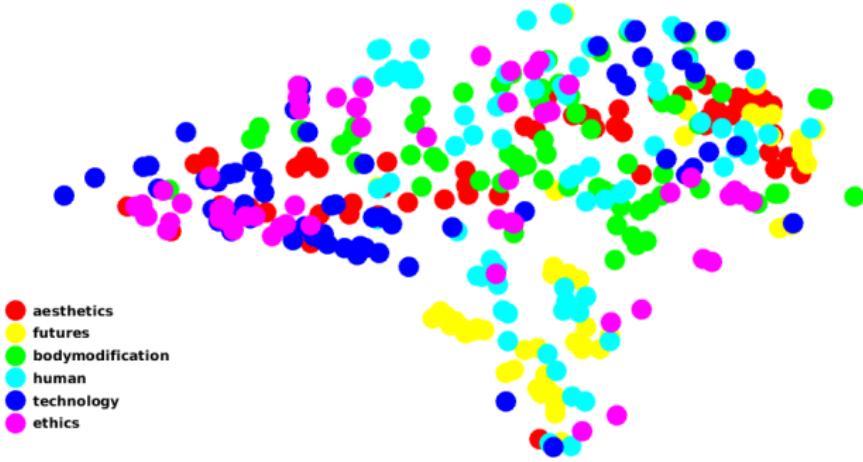


AARHUS UNIVERSITET

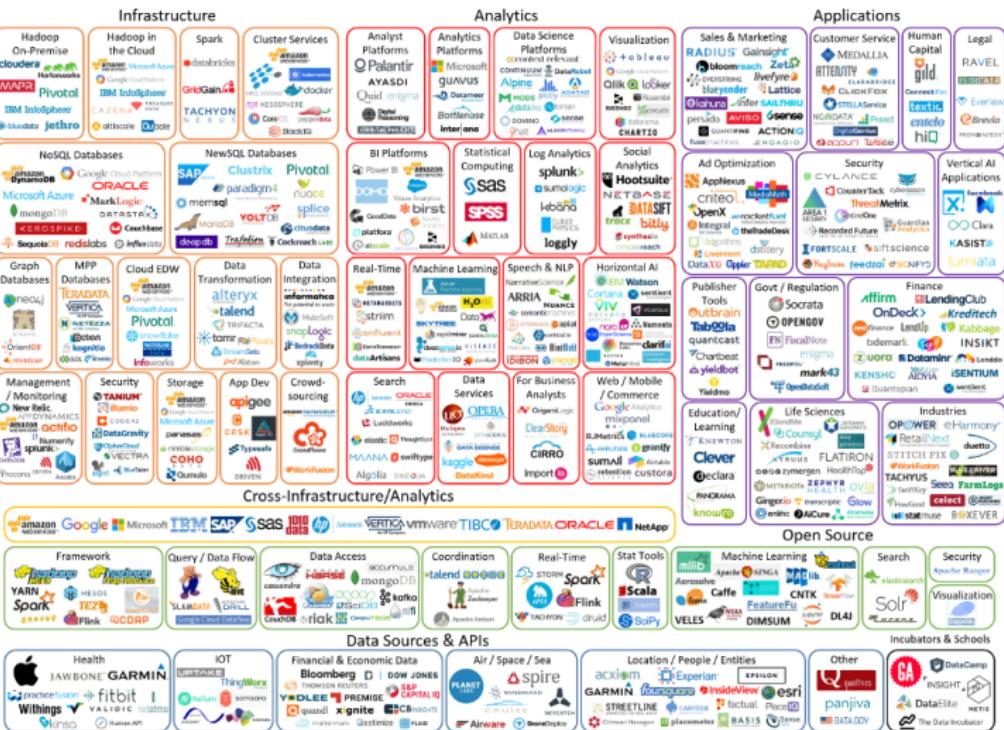
we can train an **Artificial Neural Network** to predict a word given its context (CBOW) or predict a context given a word (Skip-gram)



the word representations (word vectors) used by the ANN to solve the task will **reflect important semantic features**



## **available tools**





AARHUS UNIVERSITET

