

Outline of a Literary Informatics

Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

Dept. of History & SDU eScience Center
University of Southern Denmark

September 12, 2018

- 1 Human & Literary Informatics
 - On the role of informatics
 - A need for informatics
 - Workflows
- 2 Author dynamics
- 3 Narrative dynamics
- 4 Compositional dynamics

Human & Literary
Informatics

On the role of
informatics

A need for informatics

Workflows

Author dynamics

Narrative dynamics

Compositional
dynamics

Human & Literary
Informatics

On the role of
informatics

A need for informatics

Workflows

Author dynamics

Narrative dynamics

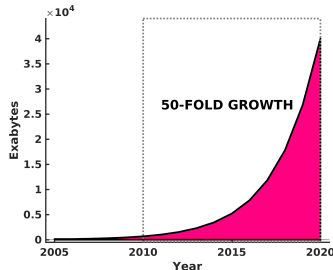
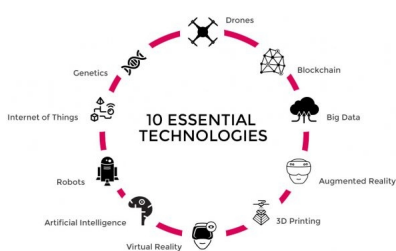
Compositional
dynamics

A need for Human & Literary Informatics

The data deluge

Outline of a Literary Informatics

Kristoffer L. Nielbo
knielbo@sdu.dk
knielbo.github.io



Human & Literary Informatics

On the role of informatics

A need for informatics Workflows

Author dynamics

Narrative dynamics

Compositional dynamics

the data deluge is transforming knowledge discovery and understanding in every domain of human inquiry

a large part of these data are soft and unstructured \Rightarrow to get value from these data, humanities (and social sciences) must utilize automation

human informatics ~ automatic information processing in the humanities

\Rightarrow *literary informatics ~ use of automation in literary studies*

Humanities need informatics

Outline of a Literary
Informatics

Kristoffer L. Nielbo
knielbo@sdu.dk
knielbo.github.io

Human & Literary
Informatics

On the role of
informatics

A need for informatics
Workflows

Author dynamics

Narrative dynamics

Compositional
dynamics



Figure 1: The increase in research publications & databases alone requires computational literacy. Publications related to Gospel of Marc (KJV) > 50K, ~ 16,500 words in 16 chp. on 11 p.

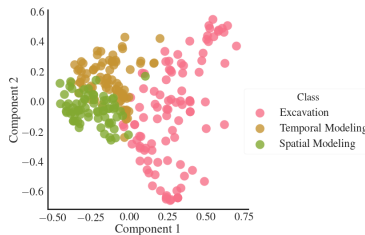


Figure 2: Advanced (human) informatics can merge, aggregate and project heterogeneous data into lower dimensional spaces that allow visual manipulation

Default workflow

Outline of a Literary
Informatics

Kristoffer L. Nielbo
knielbo@sdu.dk
knielbo.github.io

Human & Literary
Informatics

On the role of
informatics

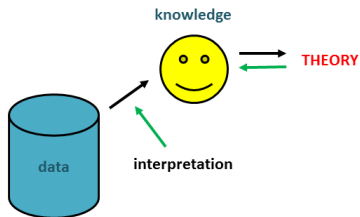
A need for informatics

Workflows

Author dynamics

Narrative dynamics

Compositional
dynamics

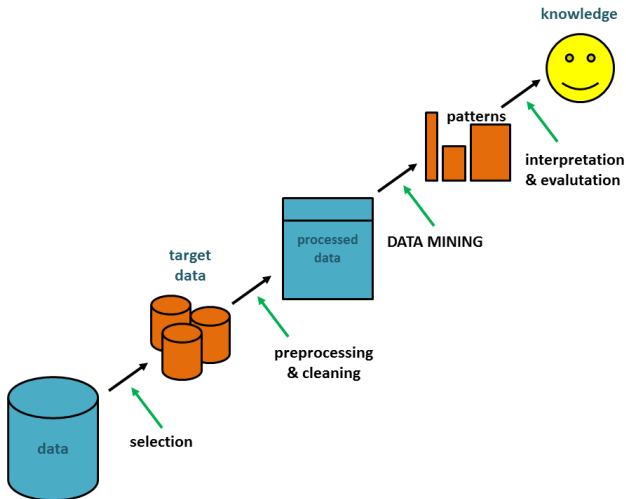


KDD workflow

Outline of a Literary Informatics

Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

Human & Literary Informatics
On the role of informatics
A need for informatics Workflows
Author dynamics
Narrative dynamics
Compositional dynamics



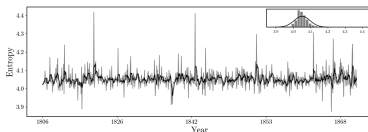
Author dynamics

The appropriate formalism

- profile the creative development of authors

- multiple measures that captures aspects of “text complexity” in terms of syllables, word and sentence length \sim characters and ngrams

- IT offers a range of “well-behaved” measures that capture lexical variability, $\sim H[X]$ as the variability of some term-vectors $X \sim$ log of the effective number of values it can take



$$\begin{aligned} H[X] &= - \sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x) \\ &= -\mathbb{E}[\log \mathbb{P}(X)] \end{aligned}$$

Competing levels of abstraction

Outline of a Literary
Informatics

Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

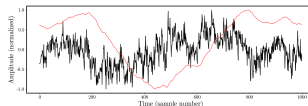
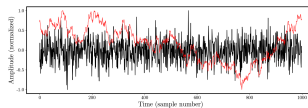
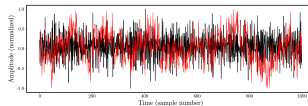
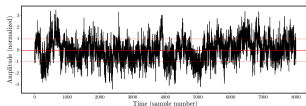
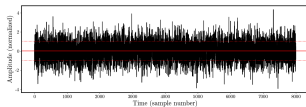
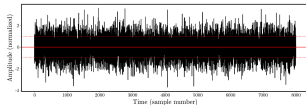
Human & Literary
Informatics

On the role of
informatics
A need for informatics
Workflows

Author dynamics

Narrative dynamics

Compositional
dynamics



$$F(X) = \left[\frac{1}{N} \sum_{i=1}^n (x_i^2) \right]^{1/2}$$

$$u(n) = \sum_{i=1}^n (x_i - \bar{x}), \quad n = 1, 2, \dots, N,$$

Noise and fractal properties

Outline of a Literary
Informatics

Kristoffer L Nielbo
knielbo@sdu.dk
knielbo.github.io

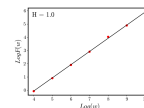
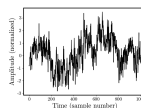
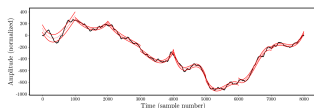
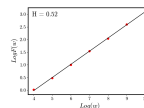
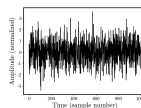
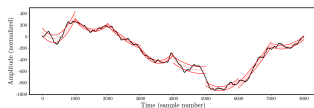
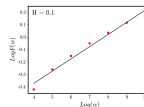
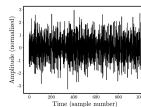
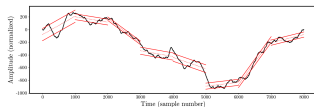
Human & Literary
Informatics

On the role of
informatics
A need for informatics
Workflows

Author dynamics

Narrative dynamics

Compositional
dynamics



Author change points

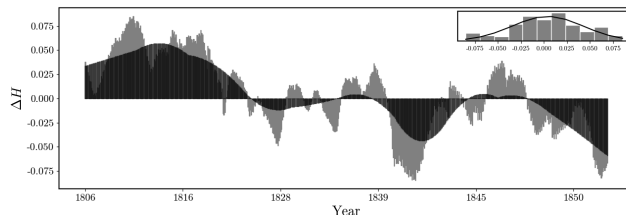


Table 1: Dominant Dynamic in the Phases of N.F.S. Grundtvig's Writings

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

Dynamic author profiling

Outline of a Literary
Informatics

Kristoffer L. Nielbo
knielbo@sdu.dk
knielbo.github.io

Human & Literary
Informatics

On the role of
informatics

A need for informatics
Workflows

Author dynamics

Narrative dynamics

Compositional
dynamics

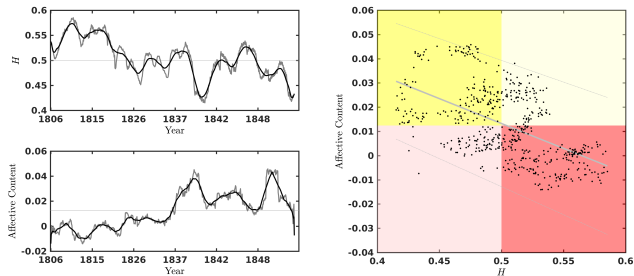


Figure 3: Combining persistent entropic trends with sentiment analysis and causal modeling, we can study **“the tormented artist”** phenomena in intellectual history.

Narrative dynamics

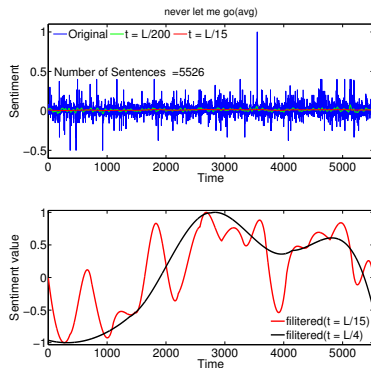


Figure 4: Story arc of Kazuo Ishiguro's 2005 novel *Never let me go*

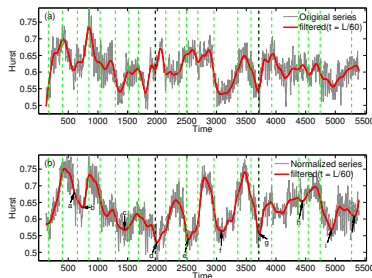


Figure 5: Evolution of the Hurst parameter under 256 window size of original and normalized sentiment time series

Compositional dynamics

Danish parsing & change detection

Outline of a Literary
Informatics

Kristoffer L. Nielbo
knielbo@sdu.dk
knielbo.github.io

Human & Literary
Informatics

On the role of
informatics
A need for informatics
Workflows

Author dynamics

Narrative dynamics

Compositional
dynamics

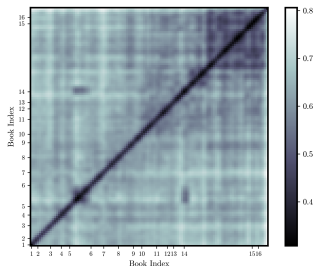


Figure 6: Cosine distance in baseline vector space model shows no evidence of change point.

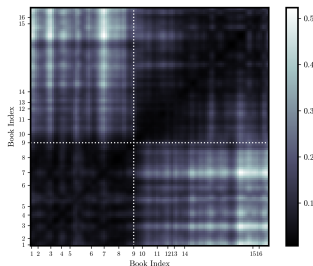


Figure 7: KL-divergence in contrast model indicates a gradual change point in book 9.

THANK YOU

knielbo@sdu.dk
knielbo.github.io

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

& credits to

Max R. Echardt and Katrine F. Baunvig, datakube, University of Southern Denmark, DK
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