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

Outline

1 Human & Literary Informatics On the role of informatics A need for informatics Workflows

2 Author dynamics

3 Narrative dynamics

4 Compositional dynamics

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

uthor dynamics

Narrative dynamics

compositional lynamics



A need for Human & Literary Informatics

Outline of a Literary Informatics

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

Human & Literary Informatics

informatics
A need for informatic
Workflows

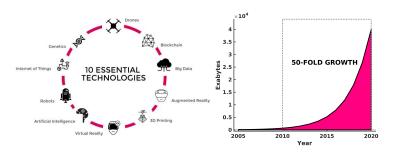
Author dynamics

Narrative dynamics

dynamics



The data deluge



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 ⇒ to get value from these data, humanities (and social sciences) must utilize automation

human informatics - automatic information processing in the humanities

 \Rightarrow literary informatics \sim use of automation in literary studies



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 informatic

Author dynamic

Narrative dynam

namics

Humanities need informatics



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

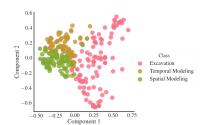


Figure 2: Advanced (human) informatics can merge, aggregate and project hetereogeneous data into lower dimensional spaces that allow visual manipulation

Outline of a Literary Informatics

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

Human & Literary Informatics
On the role of

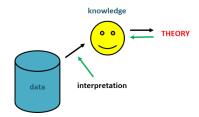
A need for informatics Workflows

Author dynamics

Composition dvnamics



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 informatic
Workflows

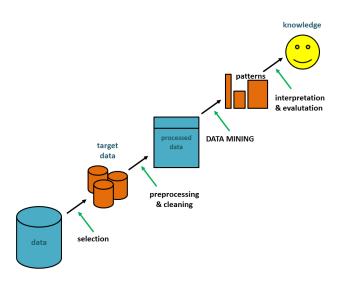
Author dynamic

Narrative dynamics

Composition



KDD workflow



Outline of a Literary Informatics

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

Human & Literary Informatics On the role of

A need for i

Author dynamics

Narrative dynamics

Compositiona



Author dynamics

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 informatic Workflows

Author dynamics

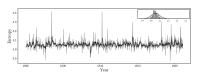
Narrative dynamics

Compositional



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



$$H[X] = -\sum_{x} \mathbb{P}(X = x) \log \mathbb{P}(X = x)$$
$$= -\mathbf{E}[\log \mathbb{P}(X)]$$

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

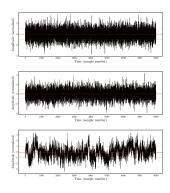
Author dynamics

Narrative dynamics

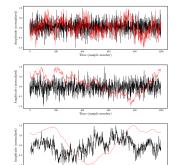
Composition dvnamics



Competing levels of abstraction



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



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

Time (sample number)

Outline of a Literary Informatics

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

Human & Literary Informatics

informatics
A need for informatics

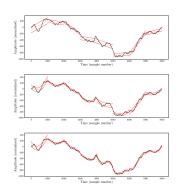
Author dynamics

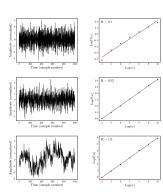
Narrative dynamics

omposition ynamics



Noise and fractal properties





Outline of a Literary Informatics

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

Human & Literary

formatics
On the role of nformatics

Author dynamics

Narrative dynamics

Compositiona Ivnamics



Author change points

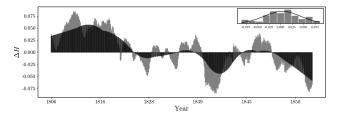


Table 1: Dominant Dynamic in the Phases of N.F.S. Grundtvig's Writings

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

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

ithor dynamics

Narrative dynamics

mpositional namics

Nielbo, K. L., K.F. Baunvig, B. Liu & Gao, J. (in print). A Curious Case of Entropic Decay: Persistent Complexity in Textual Cultural Heritage. Digital Scholarship in the Humanities.

Dynamic author profiling

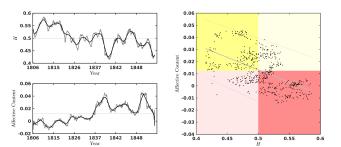


Figure 3: Combining persistent entropic trends with sentiment analysis and causal modeling, we can study "the tormented artist" phenomena in intellectual history.

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

uthor dynamics

Narrative dynamics

Composition



Narrative dynamics

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 informati Workflows

Author dynamics

Narrative dynamics

Compositional



Literature and affective computing

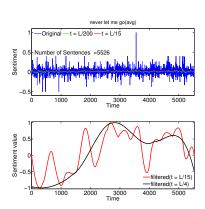


Figure 4: Story arc of Kazuo Ishiguro's 2005 novel Never let me go

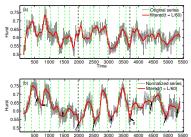


Figure 5: Evolution of the Hurst parameter under 256 window size of original and normalized sentiment time series

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 dynamic

omposition ynamics

Outline of a Literary Informatics

Hu, Q., Liu, B. Thomsen, M.R., Gao, J. & Nielbo, K.L. (*in review*). Dynamic evolution of sentiments in *Never Let Me Go*: Insights from multifractal theory and its implications for literary analysis.

Compositional dynamics

Outline of a Literary Informatics

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

Informatics
On the role of informatics
A need for informatic
Workflows

Human & Literary

Author dynamics

Narrative dynamics

Compositional dynamics



Danish parsing & change detection

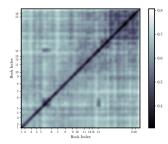


Figure 6: Cosine distance in baseline vector space model shows no evidence of change point.

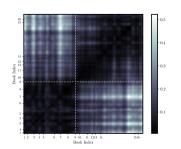


Figure 7: KL-divergence in contrast model indicates a gradual change point in book 9.

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

Lomposition Ivnamics

Nielbo, K.L., Perner, M.L., Larsen, C., Nielsen, J. & Laursen, D. (submitted). Change Detection in Gesta Danorum's Topical Composition

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

Outline of a Literary Informatics

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

Informatics
On the role of informatics
A need for informatic
Workflows

Human & Literary

Author dynamics

Narrative dynamics

Composition dvnamics

