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
   The appropriate formalism
   Competing levels of abstraction
   Noise and fractal properties
   Author change points
   Dynamic author profiling

3. Narrative dynamics
   Literature and affective computing

4. Compositional dynamics
   Danish parsing & change detection
A need for Human & Literary Informatics
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
⇒ literary informatics ~ use of automation in literary studies

Humanities need informatics

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

- **data**
- **interpretation**
- **knowledge**
- **THEORY**

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
- The appropriate formalism
- Competing levels of abstraction
- Noise and fractal properties
- Author change points
- Dynamic author profiling

Narrative dynamics
Literature and affective computing

Compositional dynamics
Danish parsing & change detection
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
The appropriate formalism
Competing levels of abstraction
Noise and fractal properties
Author change points
Dynamic author profiling

Narrative dynamics
Literature and affective computing
Compositional dynamics
Danish parsing & change detection
Author dynamics
The appropriate formalism

- profile the creative development of authors
- multiple measures that capture 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

\[
H[X] = - \sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x) = -\mathbb{E}[\log \mathbb{P}(X)]
\]
Competing levels of abstraction

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

\[ u(n) = \sum_{i=1}^{n} (x_i - \bar{x}), \quad n = 1, 2, \ldots, N, \]
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
The appropriate formalism
Competing levels of abstraction
Noise and fractal properties
Author change points
Dynamic author profiling
Narrative dynamics
Literature and affective computing
Compositional dynamics
Danish parsing & change detection

Library available at https://github.com/knielbo/saffine
Table 1: Dominant Dynamic in the Phases of N.F.S. Grundtvig’s Writings

<table>
<thead>
<tr>
<th>Time period</th>
<th>Age of onset</th>
<th>$H(X)$</th>
<th>Behavior</th>
<th>Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1806-1826</td>
<td>23</td>
<td>$H &gt; 0.5$</td>
<td>persistent</td>
<td>theoretician</td>
</tr>
<tr>
<td>1826-1839</td>
<td>43</td>
<td>$H \approx 0.5$</td>
<td>short memory</td>
<td>pragmatic</td>
</tr>
<tr>
<td>1839-1845</td>
<td>56</td>
<td>$H &lt; 0.5$</td>
<td>anti-persistent</td>
<td>breakthrough</td>
</tr>
<tr>
<td>1845-1848</td>
<td>62</td>
<td>$H \approx 0.5$</td>
<td>short memory</td>
<td>disease</td>
</tr>
<tr>
<td>1849-1872</td>
<td>65</td>
<td>$H &lt; 0.5$</td>
<td>anti-persistent</td>
<td>politician</td>
</tr>
</tbody>
</table>

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.
Narrative dynamics
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

---

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.

THANK YOU

knielbo@sdu.dk
knielbo.github.io


& 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