

# The Project Gutenberg Map

Michael Tschuggnall, Yusuf Ipek, Günther Specht

## 1 What it is

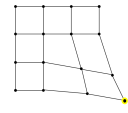
Project Gutenberg offers over 54.000 free eBooks. Using textual analysis, we computed a visual map that aligns stylistically similar texts nearby. Providing this map users can search for specific book titles or authors, or simply explore which books are similar with respect to the writing style.

## 2 Stylistic Fingerprints

- character **3-grams**
- a predefined set of other **lexical and syntactic features**, including
  - average sentence/word lengths,
  - average words/syllables per sentence,
  - punctuation ratios,
  - hapax (dis-) legomena, or
  - readability measures such as Flesch-Kincaid or Gunning Fog Index

## 3 Map Computation

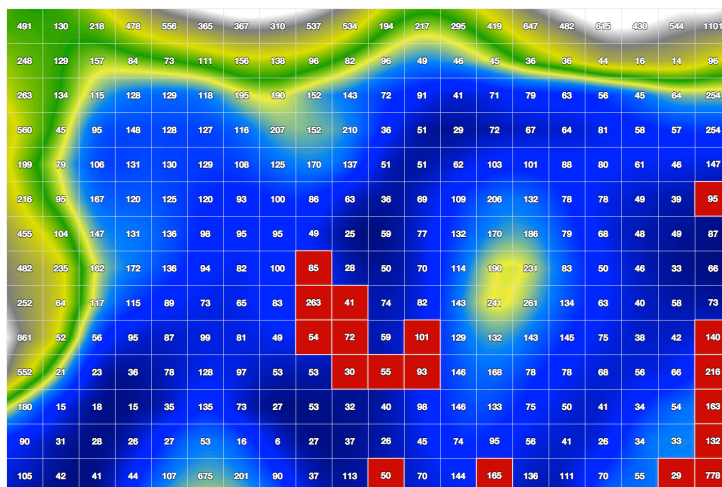
- we utilized **unsupervised learning** to build clusters using the stylistic fingerprints
- specifically, we computed a **Self-Organizing-Map (SOM)** for a fixed-sized 20x14 grid



- „Water“ with „Islands“ are created by computing **smoothed data histograms**

## 4 What you can do

- Search for specific authors or book titles** and see, which other items from the Project Gutenberg library are similar
- Choose between **different visualizations**
- Get an intuition about the **homogeneity of the writing style** of a specific author
- Detect outliers



## Example

How homogeneous are the works of Shakespeare?

## Kontakt

Michael Tschuggnall, PhD, Univ.-Prof. Dr. Günther Specht  
Department of Computer Science  
Universität Innsbruck  
<https://dbis.uibk.ac.at>