Analyzing the works of H.P. Lovecraft

Projektarbeit zum Thema Topic Modelling in Digital Humanities
von Ferris Kleier

Einführung in Digital Humanities, Wintersemester 2022/23

Autor: Ferris Kleier
Matrikelnummer: 3732130
Studiengang: Informatik B.Sc.
Datum: 14. Februar 2023

1 Introduction

This project aims to analyze the works of Howard Phillips Lovecraft, one of the most known horror authors, in regard to their themes and patterns. This will be done using digital humanities and topic modeling on all of his writings to collect a timeline-like collection of the most common themes in his works. H.P. Lovecraft highly contributed to the modern horror genre with his detailed descriptions of cosmic horror (a term defined by his writings) and stories about monsters and scary events. While his writings are known for complexity and style, his thematic and contextual depth is equally impressive and the subject of this project. Therefore, Lovecraft is well-known by fans of horror and science fiction literature, like Stephen King, who stated that Lovecraft heavily influenced his style and ideas.

Even though Lovecraft lived from 1890-1937 and did not encounter computers or other devices of our current age, his works and letters are digitalized and preserved for reading and research. One key aspect of this project is to apply the digital humanities method of topic modeling on all of Lovecraft's works. The digital humanities is a field of study, research, teaching, and invention concerned with the intersection of computing and the disciplines of the humanities. One can think of it as a bridge between the two cultures of natural science, which tries to explain what is going on, and humanities, which tries to understand what is happening. One key figure in this field is Roberto Busa (1913 - 2011), who used computational methods to create a collection of all the words of Thomas Aquinas, later known as the Index Thomisticus. He achieved his goal with the help of IBM, at that time one of the biggest computer manufacturers, and connected his studies with the help of digital computing to process the millions of words he strived to process.

By using the digital humanities, we can gain new insights on Lovecraft's writings and identify recurring themes, patterns, and motifs in the big corpus of his writings. The methods used in the digital humanities include Stylometry, Topic Modeling, Network Analysis, and Geovisualization. For the purpose of this project, we will use topic modeling to gather insight into the large corpus and structure the results like a timeline for every one of his works. Regarding digital humanities, this project aims to explore the humanities of literature and apply computational methods on it. That way we can see the most occurring themes from his first works to his last ones and compare them in terms of lore consistency (how did his themes and mythos change?), social correlation (does he refer to social events of his time?) and comprehensive insight on the different features of his

works. Some works already covered features of his writings and life like the fear of progress, dangers of scientific progress, and persistence of ancient evils.

2   Research Agenda

The research agenda of this project is 'what are the most dominant topics in H.P. Lovecraft's writings over his lifespan?'. The goal is to understand how they relate to Lovecraft's personal life experiences, historical events, cultural and social influences, as well as to the broader literary and intellectual traditions of his time. To answer that question we will apply topic modeling to each of his works in chronological order. By comparing the resulting main topics we hope to find changes in the lore of his mythos as well as correlations to the main events in his life.

We will use computational methods like text processing and tokenization to represent the main themes of Lovecraft's writings from a corpus of all of his works. Because Lovecraft died over 70 years ago, many of his writings are public domain and can easily be accessed from various institutions providing good coverage for the material. Nevertheless, some of the works he contributed are still not public domain, since some authors he worked with lived on longer than he did (e.g. Clark Ashton Smith died 1961). This is part of the reason why we decided to only cover works solely written by Lovecraft himself. The more important reason is that mixing up his writings with ghostwritten or co-written works could alter the results since we only want to cover Lovecraft's themes. Using the programming languages Python and R, we first format the corpus into a collection of all works (using Python), then process the texts to remove annotations and unnecessary titles and use topic modeling on the corpus to gather and represent the data (using R). After that, we will analyze the gathered data represented as a plotted diagram in R. That way we can create a timeline-like graphic of the main themes and motifs in Lovecraft's works and discuss them in retrospect to events or that time or his personal life.

To properly find correlations between his writings and private life or social events of his time, we will shortly examine his letters as well as his biographical records. After that, we include the ideas and insight into the discussion. We expect to find results in this aspect, since Lovecraft is known for his rejection of modern movements of that time. He was known for his dislike of New York City, after living there for a period of his life. This is one of the interesting parts of his life we hope to find reflected in his writings. We already know of the story 'The Horror at Red Hook' (1925) in which Lovecraft negatively reflects on Red Hook, Brooklyn. It is also important to mention that Lovecraft condemned ethnical groups he encountered during his time in New York, mainly Afro-Americans and Asian immigrants. It is no secret that Lovecraft used racist slurs and sometimes even antisemitic stereotypes in his writings, all of which will be analyzed according to his personal events to explore how his prejudices shaped his writing. For all of these aspects of his life, we hope to find a similar pattern in his writing through which one could pair a batch of his works to a phase of his life.

3   Data Overview

The data