

オンライン小説におけるタイトルの傾向分析

学生番号:1832028 氏名:大塚 健叶

指導教員：山崎治 准教授

1. はじめに

近年、インターネット上で個人が小説を投稿できる、利用者投稿型小説サイトが人気である。小説投稿サイトに投稿される小説に関しては、作品の画一化やタイトルの冗長性が指摘されている。これはサイト内のランキングに入るために、タイトルに利用者の目につきやすい要素を詰め込むからだと考えられている。逆説的に言えば、タイトルの情報から利用者に好まれやすい小説か判断する事が可能である。

そこで本研究では、利用者投稿型小説サイトで、利用者の評価を得やすいタイトルの傾向を明らかにすることを目的とする。このため、小説のタイトルを対象としたテキストマイニングを行い、年ごとの変化やジャンルの影響、ランキングとの関係を明らかにする。また、Web小説のタイトルを入力情報として、機械学習の手法を用いたランキングの予測を試みた。

2. オンライン小説のタイトル傾向性の調査

2.1 方法

形態素解析で小説タイトルの頻出単語の抽出と整理を行い、同一タイトルに含まれる単語の共起関係を用いた、共起ネットワーク分析を行う。また、掲載年やジャンル、ランキング(上位／中位／下位)により小説のグループ分けを行い、各グループでの頻出単語や共起ネットワークの比較を行い、傾向の違いを明らかにする。

表1に収集した小説のランキングデータの一部を示す。書籍(紙の本)とWeb小説でそれぞれ異なるサイトからランキングデータを抽出した。

表1 収集データ概要

種別	紙の本	WEB小説
収集先 サイト	ハイブリッド総合書店 「honto」	小説投稿サイト 「小説家になろう」
期間	2022年6月～9月	2022年6月～9月
データ (タイトル 数)	総合 BEST100	19ジャンル毎の BEST100

2.2 結果

収集した「小説家になろう」「honto」の小説情報メタデータを用いて、頻出単語の抽出と整理、共起ネット

ワーク分析を始めとする4手法による分析を行った。表2に、頻出単語の分析としてワードクラウドと共起ネットワークの結果を示す。

表2 紙の本とWEB小説の比較

[illegible]

分析の結果、タイトルの傾向には明らかな違いが見られた。ワードクラウド上で、紙の本では「出る」という動詞の頻度が高い一方、WEB小説では「令嬢」「最強」「世界」「追放」「転生」などの単語頻度が高く、頻出単語の種類が多かった。また、WEB小説では、緑色で表示されている形容詞の単語が多く含まれていた。

共起ネットワークから、紙の本では、小さなクラスタに分かれるが、WEB小説では、単語のつながりが大きくなることがわかった。これは「WEB小説」が「紙の本」のタイトルの文字数よりも、平均17文字程度タイトルが長く、さらにタイトルが文章になっているものが多い見られたためと考えられる。

3. おわりに

本研究では、タイトルのランキングデータに基づき、人気の高いWEB小説のタイトルに見られる傾向性の分析を行った。紙の本との比較やジャンルの違いによる分析を行い、WEB小説のタイトルにみられる特徴を明らかにすることができた。