Investigating the Transformation of Original Work by the Online Fan Fiction Community: A Case Study for Supernatural

Nina Kleindienst & Thomas Schmidt

Media Informatics Group, University of Regensburg, Germany

nina.kleindienst@stud.uni-regensburg.de
thomas.schmidt@ur.de

This proposal was accepted for presentation at the conference *Digital Practices. Reading, Writing and Evaluation on the Web.* Find more information about the conference here.

Abstract. We report upon a project investigating fan fictions for the TV show "Supernatural". Our goal is to examine if and how the fan fiction community changes the source material throughout the life cycle of the TV show. For our first analysis, we acquired two corpora: (1) all scripts of the TV show at the date of our writing and (2) over 7,000 fan fictions from the platform *Archive of Our Own*. We report first analysis focusing on the comparison of the representation frequency of the main characters of the show.

1. Introduction

We present first results of a project investigating how online fan fiction communities use the original content and deviate from it in their fan fiction stories concerning language, characters/named entities, topics and sentiments. The analysis of this transformation process has led to multiple research in the humanities (cf. Van Steenhuyse, 2011) and we want to extend this work via methods of natural language processing and text mining. We examine the use case of the popular series "Supernatural" which is among the most popular TV shows for fan fictions.

2. Corpus

We acquired a structured corpus consisting of all the scripts and metadata of all Supernatural episodes marking speakers and speeches (as of the date of this writing, the show consists of 14 season which equals to 307 episodes; see table 1).

Metric	Value
Number of seasons	14
Number of episodes	307
Number of tokens	2,182,156
Average number of tokens per episode	7,108
Average lexical diversity	0.1943

Table 1. Corpus statistics for the Supernatural TV show corpus

To create the fanfiction corpus, we scraped the popular fan fiction platform *Archive of Our Own* (AO3)[1]. For our first analysis we focused on the most popular Supernatural fan fictions (more than 300 kudos) which resulted in 7,853 works (see table 2).

Metric	Value
Number of fan fictions	7,853
Number of tokens	265,580,607
Average number of tokens per fan fiction	33,819
Average lexical diversity	0.0986

Table 2. Corpus statistics for the Supernatural fan fiction corpus

3. Analysis

Among some of our first explorations, we analyze differences in the appearances and mentions of the main characters: *Dean (Male), Sam (Male)* and *Castiel (Male)* (the latter one being introduced in season 4). Table 3 summarizes some results.

Characters	DEAN	SAM	CASTIEL (just episodes after he is introduced)
in % of episodes	99.67%	100%	46.56%
average speeches per episode	91	74	33
average percentage per episode	26.14%	21.3%	9.84%
average mentions per episode (in the speeches)	11	9	5
in % of fan fictions	89.97%	70.84%	76.03%
average mentions per fanfiction	555	209	418

Table 3. Appearances and mentions of the main characters in episodes and fan fictions

---

[1] https://archiveofourown.org/

One can easily see how the character Castiel is overrepresented in the fan fictions compared to the original work being it appearances or mentions in the text. This is in line with general findings by Milli & Bamman (2016) concerning the overrepresentation of minor characters. More analysis about the metadata of fan fictions tagged with relationships show that the tag *male/male* makes up the majority of the corpus (90%) and furthermore the relationship of *Castiel/Dean* makes up 80% of these stories (see table 4).

Universe	Alternate Universe	Canon Universe	
	37.91%	10.57%	
Rating	**Explicit**	**Explicit & Mature**	
	58.05%	73.77%	
Categories	**Female/Male**	**Female/Female**	**Male/Male**
	10.98%	2.48%	91.99%
Relationships	**Percentage**	**Relationships**	**Percentage**
Castiel / Dean	79.38% of all male/male	Jess / Sam	53.25% of all female/male
Dean / Sam	14.73% of all male/male	Ruby / Sam	7.77% of all female/male
Gabriel / Sam	9.32% of all male/male	Eileen / Sam	7.77% of all female/male
Castiel / Sam	1.8% of all male/male	Castiel / Meg	6.15% of all female/male

Table 4. Distributions of selected fan fiction metadata

This phenomenon concerning the dominance of homo-romantic content in fan fictions has been already identified in the humanities (cf. Hellekson & Busse, 2006; Tosenberger, 2008) and can be seen here via quantitative methods. Overall, the results also show how the online community influences the original work since the online popularity of Castiel lead to him becoming a major character in the show.

4. Future Work

While our findings are currently limited on the use case of Supernatural, we are able to gain insights on how the fandom online community transforms the source material and differentiates from it via computational methods. We plan to further our analysis by applying sentiment analysis and topic modeling on our corpora. We would be happy to present these and more results as a poster, which also offers us the possibilities to discuss our plans and current problems.

References

Hellekson, K. & Busse, K. (2006). *Fan Fiction and Fan Communities in the Age of the Internet: New Essays.* Jefferson, NC: McFarland.

Milli, S., & Bamman, D. (2016, November). Beyond canonical texts: A computational analysis of fan fiction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 2048-2053).

Tosenberger, C. (2008). Homosexuality at the online Hogwarts: Harry Potter slash fan fiction. *Children's Literature*, 36(1), 185-207.

Van Steenhuyse, V. (2011). The writing and reading of fan fiction and transformation theory. *CLCWeb: Comparative Literature and Culture*, 13(4), 4.