

## **Anna Moskvina, “Genre Identification with Graph Neural Networks”**

**Abstract:** This paper presents a study on automatic genre identification from short texts (blurbs). We use Graph Neural Networks (GNN) and provide a detailed error analysis of the classification results provided by the trained model. Automatic genre identification and prediction is not a new task in the Digital Humanities (Worsham, Kalita, 2018; Calvo Tello, 2018), as genre classification schemes for fiction and non-fiction are numerous and well-established. There has been a multitude of research that takes into account the textual information from the books paired with, for example, emotional analysis (Kim, 2017; Samothrakis, Fasli, 2015). Most of these systems work with a well-defined, not overlapping set of genres. Nevertheless, the researchers with more humanities background strongly believe that genres that we observe as readers are less structural. The paper “The life cycles of genres” (Underwood, 2016) claims that what we usually assume as genres, may be better called „entities of different kinds“. Antoniak et al (2021) in their study showed that genres, when books are labelled by readers themselves, do overlap.

One of the genre classification schemes that reflects the overlapping nature of different genres was introduced as a part of GermEval Task on the classification of German blurbs in 2019 (Remus et al, 2019), where the genres were represented as a hierarchy with multiple layers and several genres could be assigned to one book. Each book was represented as a short description or a “blurb”. For example, one blurb could receive the following labels (different layers of the hierarchy are connected in this example with a “-“): “Literature and Entertainment-Fantasy-Adventure Fantasy“ and „Literature and Entertainment-Fantasy- heroic Fantasy“.

For our research we use the textual information from the “blurb” and additionally the information of how the genres relate to each other, by implementing a Graph Neural Network (Zhang et al, 2020). This way we can build a system that can automatically predict what genres a book might be assigned to. Within our paper we delve more into qualitative error analysis with the intent to better understand which errors our system has made (whether the genres were not well-defined, or too detailed, or too vague) and why as well as to see if the errors made by an intelligent system can reflect on the given classification scheme (Remus et al, 2019).

**Anna Moskvina** studied theoretical and computational linguistics at Moscow State Linguistic University (2007-2012) and computational linguistics at the University of Stuttgart (2014-2017). Since 2017, she has been working as a research assistant and a PhD student at the University of Hildesheim. From 2017 to 2020, she worked in the research project Rez@Kultur, where she was primary responsible for collecting and processing research data. She has been a member of the German Society for Computational Linguistics and Language Technology (GSCL) since 2020. Her thesis focuses on the use of deep learning methods in the Digital Humanities.