

Srishti Sharma, Independent Researcher (srishti.0118@gmail.com)
Report on the Computational Literary Studies Infra Transnational Access Fellowship at
the Ghent Center of Digital Humanities, Ghent University
(Ghent, Belgium 25 April – 17 July 2022)

Project title: Can a book make you happy? Predicting emotional links between genre, plot and reader response

Introduction

My purpose for submitting an application to the CLS Infra TNA Fellowship was primarily to be part of the digital literary studies community and garner the technical expertise necessary for the advancement of said research. My project proposal focused on the evaluation of the emotional impact of fiction on readers through the lens of their corresponding reviews. I wanted to assess whether the emotion represented and expressed in fiction can predict the emotions perceived and expressed by readers in their reviews. For instance, do tragic stories prompt more negative wordings in reviews? This particular research project required familiarity with literary text analysis especially sentiment analysis techniques which I found at the Ghent Center for Digital Humanities, Ghent University. The team at Ghent CDH, particularly Dr. Julie Birkholz and PhD student Tess Dejaghere, was proficient in all of the above-mentioned areas of research and offered an environment with high potential for growth. They also worked closely with the Language Technologies Group, LT-3, at Ghent University who share their computational linguistics expertise.

Aims of the Project

The goals of my fellowship included gaining access to and learning more about computational infrastructure as well as social infrastructure vital to digital humanities research. This involved trying out various computational methods to find the one best suited for my emotion detection research on literary texts. I was also invited to be a part of the digital humanities community by participating and engaging in insightful discussions at cross-border conferences and events. The main aim of the research visit was to receive research guidance and computational support to successfully complete the sentiment analysis and emotion detection of the chosen corpus using state-of-the-art techniques to achieve the best possible result. I was also exposed to the sociocultural aspect of being part of a research team and exchanging thoughts and ideas on a regular basis for the overall development of the field.

Description of the Research Visit

My Research Visit was a combination of technical, social and cultural growth. The team at Ghent CDH and CLS Infra TNA Fellowship had planned a well-balanced visit where the majority of my time was devoted to learning and experimenting with new computational models fit for sentiment analysis of English novels and their Goodreads reviews. It involved a step-by-step

process starting from data mining of Goodreads reviews of obscure titles to processing the extracted text data and making it fit to input it into the various models. It is as follows:

- Once the data collection of around 450 English novels and close to 1 million English reviews was done, the processing steps were laid out.
- We made sure that every sentence in the books dataset and the review dataset was written in English, and any unwanted text was manually removed.
- We also experimented with the tokenization of books based on the page, paragraph and sentence level tokens and compared their performance.
- Additionally, character masking was also suggested by previous research and we carried it out to see if the model would automatically associate positive or negative sentiment with certain character names.
- This is usually observed and can be solved by masking the character or author names from the dataset, however in our case masking caused the performance to dip. Hence we used the corpus without masking the character names.

Post the processing steps, a number of transformer-based models were tested with the dataset to such as distilBERT, GoEmotion, XLM base model and so on. The best-performing model turned out to be the fine-tuned XLM RoBERTa model. The model was fine-tuned separately for the books and the reviews dataset. For the books corpus, the [DENS](#) dataset was used for fine-tuning. For the reviews dataset, we used a model shared by Alessandro Fossati, University of Milan-Bicocca. The model was trained on annotated social media and Reddit comments as well as movie reviews. Furthermore, two emotions were excluded from the analysis viz. Love and disgust, due to poor performance while training. This can be attributed mostly to the lack of data available on texts annotated with the “disgust” tag and the overlap of “love” and “joy” over most instances in the annotated dataset. The final predictions of sentiment and emotion detection were calculated genre-wise. The F1-score for the task of sentiment analysis of the books dataset by the selected model turned out to be 0.745. These predictions will then be further used to calculate the correlation between the sentiment of the books and the sentiment of the corresponding reviews for each genre. All of these insights were finally presented by me during a [talk](#) towards the end of my visit. The CLS Infra TNA Fellowship team, Ciara Murphy, also conducted an [interview](#) before the talk to discuss the project and the motivation behind it. I was surprised to hear the positive feedback and interest shown by the researchers invited by the Ghent CDH team as part of a Library Lunch event.

While working on my research, I also had the wonderful opportunity to meet a number of researchers within the field, sharing similar or interesting ideas. This, I believe, had a huge impact on my understanding of the field and expanded my scope of knowledge. I was fortunate to attend insightful talks conducted during the Literature and Computation Week at UC Louvain, DH Lecture Series at University of Antwerp and conferences such as DH Benelux held at University of Luxembourg and the Computational Stylistics Workshop on Emotion and Sentiment Analysis in Literature, Paris. I was amazed to hear about all the research being done in digital humanities and the intersection at which it stands, sparking interest from researchers across domains. I met people from the History Department as well English and German Literature department which resulted in fruitful discussions revolving around the methodology

used for analysis of reader response, digital tools literacy and the intricacies of analysing the visual-focused children's literature and comic books.

Concluding remarks

The research visit was extremely successful in terms of gaining technical expertise and expanding my social circle of relevant researchers. I was fortunate to have met many incredible and highly motivated individuals throughout my visit. It definitely poised me for a better future and introduced me to all the possible opportunities in the field. Overall I received an introduction on the various methods of conducting literary analysis as well as the benefits and limitations of computational methods in the study of emotions. I realized that working with literary texts is complex and requires frameworks that account for the vast variety of styles and languages in literature. The experience was highly rewarding and I would recommend it anyone looking for a similar opportunity.

Acknowledgements

I would like to express my gratitude to the CLS Infra TNA fellowship team, especially Marco Raciti, for presenting such a learning opportunity and to the GhentCDH team for inviting me and providing an environment for excellence and growth. The team, especially Julie Birkholz, Tess Dejaghere and Lise Foket, was more than welcoming and made me feel at home. Moreover, I'd like to thank Christophe Verbruggen and Leen Verhoeve for helping out with all the administrative procedures for making this visit possible. I was inspired by the commitment and enthusiasm shown by the researchers I met during my stay such as Sally Chambers, Ray Siemens and many others. I am grateful to Els Lefever and Pranaydeep Singh, from the LT-3 team, who showed interest and also supported my research while at Ghent University. I am confident that the experience will add immense value to my future research endeavours.