

CLEF 2023: Scientific Text Simplification and General Audience

Notebook for the SimpleText Lab at CLEF 2023

Felix Ohnesorge¹, Mari Ángeles Gutiérrez² and Julia Plichta³

1 University of Kiel, 24118 Kiel, Germany

2 University of Cádiz, 11003, Cádiz, Spain

3 University of Gdańsk, 80-309 Gdańsk, Poland

Abstract

This project aimed to enhance comprehension of scientific texts for a general audience through two tasks: Complexity Spotting and Text Simplification. Complexity Spotting involved ranking and explaining difficult concepts using the SimpleT5 model, overcoming challenges like lack of feedback and time constraints. Text Simplification utilized SimpleT5 and GPT-3 models to provide simplified versions, facing challenges such as time limitations, token usage, and limited GPU resources. Evaluation showed SimpleT5 outperforming GPT-3 in key metrics. Finally, the experiment demonstrated the effectiveness of models and potential for future advancements in simplifying scientific text.

Keywords

Natural Language Processing, Scientific Information Access, Text Simplification, Sentence Transformers

¹ CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece

EMAIL: felix-eutin@gmx.de (Felix Ohnesorge), mariangeles.gutierrez@uca.es (Mari Angeles Gutierrez), j.plichta.162@studms.ug.edu.pl (Julia Plichta)

© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Task 2: Complexity Spotting: Identifying and Explaining Difficult Concepts for a General Audience

1. Introduction

The goal of these two tasks of the SimpleText lab, organized as part of the Clef 2023 conference [1], was, in this case, to identify and explain difficult concepts in a scientific text to assist readers in understanding complex information. Participants were required to rank the difficulty of up to five terms and provide short explanations for each identified difficult term. The provided passages are treated as independent sentences, allowing for repetition of difficult terms.

2. Tasks performed

The participant used the SimpleT5 model for this task. The workflow followed the following steps:

- a. Train Dataset Preparation – created a train dataset with columns "source_text" and "target_text" and added the prefix "identify difficult term: " to each source_text.
- b. Model Training – the pretrained SimpleT5 model was trained on the train dataset for five epochs. The best model was selected based on the performance after the second epoch, with a train loss of 0.5293 and a validation loss of 1.324.
- c. Test Dataset Preparation – prepared a large test dataset and added the prefix "identify difficult term: " to each source_text.
- d. Predictions – to enhance efficiency, the large dataset was split into 100 chunks. Partial predictions were performed, saved in JSON files, and subsequently joined together.

3. Challenges Faced

The participant faced the following challenges during the task:

- a. Lack of feedback – Unfortunately, the participant did not receive any feedback on the submitted results, making it difficult to assess the effectiveness of the approach.
- b. Time constraints – Due to time limitations, the participant was unable to explore alternative models or complete task 2.2 within the given timeframe.
- c. Time-intensive predictions – Generating predictions for the large dataset took a significant amount of time. For each sentence, the model extracted only one difficult term.

4. Conclusion

In conclusion, the participant made an attempt to tackle the complexity spotting task using the SimpleT5 model. Although the workflow was successfully implemented, the lack of feedback and time constraints hindered further analysis and improvement. Additionally, the time-intensive nature of predictions for the large dataset limited the model's output to only one difficult term per sentence. Future work could involve evaluating alternative models, addressing the challenges faced, and enhancing the performance and efficiency of the complexity spotting task.

Task 3: Text Simplification: Scientific Text Simplification

1. Introduction

The task focuses on providing simplified versions of text passages, specifically popular science articles and queries matched with abstracts of scientific papers. The evaluation of the simplifications includes assessing complexity, errors, and information distortion [2]. Text complexity and simplification have been extensively studied in linguistics, education science, and natural language processing. Simplified texts are valuable for non-native speakers, young readers, individuals with reading disabilities, and those with lower levels of education. As Ermakova *et al.* refers [3], text complexity and simplification have been extensively studied in linguistics, education science, and natural language processing. Simplified texts are valuable for non-native speakers, young readers, individuals with reading disabilities, and those with lower levels of education. In 2023, the training and evaluation data have been expanded, and large-scale automatic evaluation measures, along with small-scale human evaluation, are utilized; however, defining the desired output of simplification remains a challenge, as traditional readability scores only consider word and sentence length, while vocabulary-based metrics overlook information distortion.

2. Tasks performed

The participant employed two models, SimpleT5 and GPT-3, to perform the text simplification task. The workflow followed these steps:

- a. Prepared a train dataset by merging input and output sentences.
- b. Added the prefix "simplify: " to each source_text.

For SimpleT5:

- c. Trained a pretrained SimpleT5 model on the train dataset for five epochs. The best model was selected based on the performance after the first epoch, with a train loss of 0.9715 and a validation loss of 1.1263.
- d. Predicted simplified sentences using the trained model on a small test dataset.

For GPT-3:

- e. Split the test dataset into 20 chunks for efficiency.
- f. Performed partial predictions using GPT-3 and saved the results in JSON files.
- g. Joined all JSON files to compile the predictions.

Note: The small dataset was chosen to reduce evaluation time.

3. Challenges Faced

The task of scientific text simplification encountered several challenges, including:

- a. Time limitations: The project faced time constraints, which affected various aspects of the experiment, such as dataset preparation, model training, and evaluation. The limited timeframe required balancing efficiency and accuracy in the implementation process.
- b. Time-intensive predictions: Predicting simplified sentences using models like SimpleT5 and GPT-3 can be time-consuming, especially when dealing with large datasets. Generating predictions for each source sentence from the test dataset proved to be a time-intensive task, resulting in potential delays and longer evaluation times.
- c. Costly predictions in terms of tokens: Text simplification tasks often involve working with large amounts of text, which requires a significant number of tokens. Models like GPT-3 have token limits, and exceeding these limits incurs additional costs. Managing the token usage and optimizing the predictions within the token constraints were essential to avoid excessive expenses.
- d. Limited GPU resources in Google Colab: Google Colab, a widely used platform for training and inference, provides limited access to GPUs. This limitation can impact the efficiency and speed of the prediction process, particularly when dealing with resource-intensive models like GPT-3. Managing and optimizing GPU resources was crucial to mitigate the challenges posed by limited availability.

4. Evaluation Results

The evaluation results for both models, GPT-3 and SimpleT5, are as follows:

- a. GPT-3:
FKGL: 8.083, SARI: 34.597, BLEU: 6.909, Compression ratio: 0.433, Sentence splits: 0.989, Levenshtein similarity: 0.482, Exact copies: 0.0, Additions proportion: 0.158, Deletions proportion: 0.715, Lexical complexity score: 8.620.
- b. SimpleT5:
FKGL: 13.397, SARI: 41.401, BLEU: 45.189, Compression ratio: 0.894, Sentence splits: 0.996, Levenshtein similarity: 0.915, Exact copies: 0.002, Additions proportion: 0.073, Deletions proportion: 0.172, Lexical complexity score: 8.686.

5. Analysis of the Results

Based on the evaluation metrics, it can be observed that SimpleT5 outperformed GPT-3 in terms of SARI, BLEU, compression ratio, sentence splits, Levenshtein similarity, and lexical complexity score. SimpleT5 achieved higher scores in most metrics, indicating better quality in terms of similarity to the reference simplifications, concise output, preserved sentence boundaries, and lower lexical complexity. However, SimpleT5 had a higher FKGL score, suggesting slightly more complex text compared to GPT-3.

6. Conclusion

In conclusion, the task of scientific text simplification presented various challenges, including time limitations, time-intensive predictions, and costly predictions in terms of tokens and limited GPU resources. Despite these challenges, the experiment effectively showcased the effectiveness of models such as SimpleT5 and GPT-3 in simplifying scientific text. Particularly, the SimpleT5 model demonstrated superior performance compared to GPT-3, as evidenced by its higher scores in evaluation metrics. This indicates that SimpleT5 provided better results in terms of generating simplified versions of scientific text passages. The findings highlight the potential of advanced models like SimpleT5 for achieving high-quality simplifications in the scientific domain. To further enhance the task, future work can explore alternative models and techniques to improve efficiency and address the challenges faced. Additionally, considering the use of alternative platforms or cloud-based solutions with increased GPU resources could help mitigate the limitations encountered. By addressing these challenges and exploring new avenues, the field of scientific text simplification can continue to advance and provide valuable simplified versions of complex scientific content.

7. References

- [1] Liana Ermakova, Eric SanJuan, Stéphane Huet, Olivier Augereau, Hosein Azarbondy, and Jaap Kamps. 2023. Overview of SimpleText - CLEF-2023 track on Automatic Simplification of Scientific Texts. In Avi Arampatzis, Evangelos Kanoulas, Theodora Tsikrika, Stefanos Vrochidis, Anastasia Giachanou, Dan Li, Mohammad Aliannejadi, Michalis Vlachos, Guglielmo Faggioli, Nicola Ferro (Eds.) *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023).

- [2] M. Louis, S. Humeau, P.-E. Mazaré, A. Bordes, É. V. de La Clergerie, and B. Sagot, Reference-Less Quality Estimation of Text Simplification Systems, arXiv (2019). <http://arxiv.org/abs/1901.10746>.
- [3] Liana Ermakova, Eric SanJuan, Stéphane Huet, Olivier Augereau, Hosein Azarbondy, and Jaap Kamps. 2023. CLEF 2023 SimpleText Track: What Happens if General Users Search Scientific Texts?. 10.1007/978-3-031-28241-6_62.