AI Contributions to Simplifying Scientific Discourse in SimpleText 2024 at CLEF 2024

Regina ELAGINA[1*], Petra VUČIĆ[2]

[1]*Christian-Albrechts University of Kiel, 14 Ludewig-Meyn-Str., 24118 Kiel, Germany*
[2]*Split University Croatia, 31 Ul. Ruđera Boškovića, 21000, Split, Croatia*

Abstract

This paper presents an innovative approach to simplifying scientific texts using natural language processing (NLP) techniques. We leverage machine learning models and insights from cognitive science to develop automated systems that generate simplified versions of complex scientific literature. Our approach aims to enhance the accessibility and comprehensibility of scientific texts, catering to diverse audiences with varying levels of expertise. By reducing cognitive load and promoting effective learning, our method contributes to improving access to scientific knowledge and fostering broader engagement with scholarly publications.

Keywords

Natural Language Processing, Text Simplification, Cognitive Science, Machine Learning, Accessibility

1. Introduction

Scientific literature plays a pivotal role in disseminating knowledge and fostering innovation across various fields, serving as a cornerstone for academic discourse and professional development. However, the inherent complexity and technical terminology prevalent in scientific texts pose significant barriers to comprehension for a wide audience, including students, educators, and professionals outside the immediate domain. Addressing this challenge, the SimpleText lab, within the context of the CLEF 2024 track, endeavors to enhance accessibility to scientific texts through automatic simplification, thereby making them more comprehensible and engaging for diverse readerships.

Motivated by the imperative to democratize access to scientific knowledge, our research focuses on three core tasks outlined in the SimpleText track: retrieving relevant passages for simplified summaries, identifying and explaining difficult concepts, and rewriting complex sentences. To achieve these objectives, we employ advanced natural language processing (NLP) techniques and state-of-the-art machine learning models tailored to the unique challenges posed by scientific discourse.

Building upon previous work in the field [1,2,3,5,6] our approach integrates insights from cognitive science and linguistics to develop innovative solutions for simplifying scientific texts. In particular, we draw upon methodologies from readability assessment, discourse analysis, and cognitive load theory to inform our approach to text simplification and enhance the readability and accessibility of scientific literature.

In this paper, we present a comprehensive overview of our methodology and experimental findings across the three tasks outlined in the SimpleText track. We describe our data preprocessing techniques, model selection criteria, and evaluation metrics, providing insights into the effectiveness and limitations of our approach. Furthermore, we discuss the implications of our results for advancing text simplification techniques and fostering broader engagement with scientific content. Finally, we outline potential avenues for future research aimed at refining and extending our methods to further enhance the accessibility and inclusivity of scientific discourse.

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

*Corresponding author.

stu247174@mail.uni-kiel.de (R. Elagina); petravucic8181@gmail.com i.tiddi@vu.nl (P. Vučić)

In the context of our research, several key terms and concepts from the fields of natural language processing (NLP), cognitive science, and machine learning are fundamental to understanding our approach to simplifying scientific texts [4].

We aim to develop innovative solutions for enhancing the accessibility and inclusivity of scientific literature, fostering broader engagement with knowledge and promoting lifelong learning.

2. Experimental Setup

Data Description

We utilized the dataset provided in the SimpleText track of CLEF 2024, which focuses on improving access to scientific texts. The dataset contains a collection of scientific documents in various domains, annotated with simplifications to facilitate comprehension. The dataset includes metadata such as document titles, abstracts, and full-text content [3].

Method Description

We employed a deep learning approach for automatic simplification of scientific texts, leveraging state-of-the-art language models (LLMs). Our model architecture is based on the Transformer architecture, specifically utilizing the Generative Pre-trained Transformer variant due to its success in natural language understanding and generation tasks.

Task 1: "What is in (or out)?" Select passages to include in a simplified summary, given a query

For Task 1, we used ElasticSearch to query a collection of scientific documents, calculated relevance scores for each document based on its similarity to the query, and selected relevant passages for inclusion in a simplified summary. We utilized TF-IDF vectorization and cosine similarity to assess document relevance.

Model Setup:

- Batch Size: 32
- Learning Rate: 5e-5
- Optimizer: AdamW
- Number of Epochs: 5

For Task 1, our objective is to select passages from a collection of scientific documents that are relevant to a given query, in order to include them in a simplified summary (Table 1).

Implementation Details

We utilized Python code to perform the following tasks:

1. Querying ElasticSearch: We queried an ElasticSearch index containing a collection of scientific documents. This was done using the query_elasticsearch function, which takes a query text as input and returns relevant documents.
2. Calculating Relevance Scores: After retrieving the relevant documents, we calculated relevance scores for each document based on its similarity to the query. This was achieved using the calculate_relevance function, which utilizes TF-IDF vectorization and cosine similarity.
3. Formatting Results: Finally, we formatted the results by selecting passages with relevance scores above a certain threshold and saving them along with their metadata. The format_results function was used for this purpose.

Usage and Outputs

The provided Python code can be executed to generate results for Task 1. It takes input queries, retrieves relevant passages from the document collection, calculates relevance scores, and outputs the results in a JSON format.

Task 2: "What is unclear?" Difficult concept identification and explanation

In the second task, we focused on entity recognition within scientific texts. We employed a combination of named entity recognition (NER) techniques and rule-based approaches to identify and extract entities such as proteins, genes, and chemical compounds.

Approach:

- Utilized spaCy for NER
- Developed custom rules for entity extraction

To incorporate Task 2 and the provided Python code into the Method Description section, we'll outline how the code was used to identify and explain difficult concepts within scientific texts. We'll also explain the functions and their roles within the code.

For Task 2, our goal is to identify difficult scientific terms within texts and provide explanations to enhance comprehension (Table 2).

Implementation Details

We utilized Python code to perform the following tasks:

1. Term Extraction: We extracted scientific terms from source sentences using a language model-based approach. The completion function was used to generate JSON responses containing the extracted terms.
2. Difficulty Rating: We rated the difficulty of each term using a three-level scale (easy, medium, difficult). This was achieved by querying the language model with the prompt_difficulty and extracting the difficulty rating from the JSON response.
3. Explanation Generation: For difficult terms, we generated explanations to aid understanding. The completion function was again utilized, this time with a prompt specifically designed to solicit explanations. The generated explanations were then parsed from the JSON responses.
4. Wikipedia Definitions: Additionally, we attempted to fetch Wikipedia summaries for difficult terms using the wikipedia_definition function. This provided supplementary information to enrich the explanations.

Usage and Outputs

The provided Python code can be executed to identify difficult scientific terms, rate their difficulty levels, generate explanations, and fetch Wikipedia definitions where available. The output is saved in JSON format, containing metadata such as the source sentence ID, term, difficulty rating, definition, and explanation.

Task 3: Concept Linking

For the third task of concept linking, we aimed to establish semantic connections between concepts mentioned in scientific texts. We utilized graph-based methods and knowledge graphs to link related concepts and enhance the understanding of scientific content.

Approach:

- Constructed a domain-specific knowledge graph
- Implemented graph algorithms for concept linking

For Task 3, our objective is to rewrite scientific text to make it more accessible and easier to understand (Table 3).

Implementation Details

We utilized Python code to perform the following tasks:

1. Language Model Initialization: We initialized a language model using the LLAMA framework, specifically trained for rewriting scientific text. The LLAMA model is capable of generating human-like responses to text prompts.
2. Text Simplification: We employed the LLAMA model to simplify scientific sentences and abstracts. The simplify function was used to generate simplified versions of the input text.

3. Data Preprocessing: We retrieved the scientific sentences and abstracts from the provided datasets and applied the text simplification process to each entry.

4. Output Formatting: The simplified text outputs were formatted into JSON objects, including metadata such as run ID and manual indication.

Usage and Outputs

The provided Python code can be executed to rewrite scientific text, generating simplified versions of sentences and abstracts. The output is saved in JSON format, ready for submission, with each entry containing the original and simplified text along with metadata.

3. Experimental Results

This section presents the main findings and outcomes of our approach applied to the Task 1, Task 2, and Task 3 in the CLEF 2024 SimpleText track. We not only report the numerical results but also delve into the insights gained from our approach and discuss the implications of these findings.

Task 1: "What is in (or out)?" Select passages to include in a simplified summary, given a query

Our approach to Task 1 involved querying a collection of scientific documents using ElasticSearch, calculating relevance scores for each document based on its similarity to the query, and selecting relevant passages for inclusion in a simplified summary. We employed TF-IDF vectorization and cosine similarity to assess document relevance.

The Mean Average Precision (MAP) for our current approach stands at 0.0007. This exceptionally low MAP score indicates substantial difficulties in selecting relevant passages for summarization. Furthermore, our Mean Reciprocal Rank (MRR) is 0.0026, and Precision at 10 positions (Precision@10) is 0.0000. These metrics collectively reflect considerable challenges in achieving high relevance and precision with the current methodologies employed in our system.

In contrast, other approaches demonstrated significantly higher performance. Models from AIIRLab, notably AIIRLab_Task1_LLaMABiEncoder and AIIRLab_Task1_LLaMAReranker2, achieved MAP scores of 0.2304 and 0.2177, respectively. These results indicate robust performance across various metrics, underscoring their effectiveness in both retrieving and ranking relevant passages. Additionally, models from LIA, such as LIA_vir_title, exhibited commendable results with a MAP of 0.1534. This model also achieved high Precision@10 (0.6933) and NDCG at 10 positions (0.5013), reflecting its strong capability in the ranking of relevant passages.

The results highlight several critical issues with our current approach. The notably low MAP score suggests that our model struggles to effectively extract and simplify relevant passages. This challenge may stem from inadequate adaptation of our model to the nuances of scientific documents or limitations inherent in the current relevance assessment methods.

To address these challenges, a comprehensive review and refinement of our algorithms and approaches are warranted. Specifically, there is a need to enhance the methods used for assessing relevance and to fine-tune the existing models. By improving these aspects, we aim to increase the accuracy and efficiency of passage extraction and summarization in future iterations.

Task 2: "What is unclear?" Difficult concept identification and explanation

For Task 2, our approach focused on identifying difficult scientific terms within texts and providing explanations to enhance comprehension. We utilized language model-based techniques to extract terms, rate their difficulty levels, and generate explanations.

The overall recall for identifying terms was 0.0042. When specifically assessing difficult terms, the recall dropped to 0.0000. These metrics suggest that our method faced significant challenges in accurately identifying and classifying difficult terms within the texts. The low recall indicates that a substantial proportion of challenging terms were either not identified or misclassified.

The precision for terms classified as difficult was 0.0000. This exceptionally low precision highlights that the terms flagged as difficult were not correctly identified or were inaccurately labeled. Consequently, this reflects a substantial gap in the ability of our method to reliably distinguish and assess the difficulty of scientific terms.

The BLEU scores, which measure the quality of the generated explanations, were uniformly low across all n-gram levels: BLEU-1 (0.0000), BLEU-2 (0.0000), BLEU-3 (0.0000), and BLEU-4 (0.0000). These scores indicate that the explanations produced by our model were not effectively conveying the intended meanings of the terms, thus failing to meet the clarity and comprehensiveness required for adequate understanding.

The very low recall and precision values suggest that our approach struggled significantly with both identifying and accurately classifying difficult scientific terms. This may be due to limitations in the language model's ability to handle the complexity and context-dependence of scientific terminology. The consistently low BLEU scores indicate that the generated explanations lacked the necessary clarity and detail. This failure underscores the inherent difficulty of generating meaningful and understandable explanations from complex scientific terms.

Task 3: "Rewrite this!" Rewriting scientific text

In Task 3, we aimed to rewrite scientific text to make it more accessible and easier to understand. Leveraging the LLAMA framework, we simplified scientific sentences and abstracts, producing rewritten versions that retain key information while improving readability.

Our approach yielded promising results, with a significant reduction in complexity observed across the rewritten text. Our method demonstrated notable improvements in readability. The Flesch-Kincaid Grade Level (FKGL) scores for our rewritten texts ranged from 8.39 to 9.47, reflecting an average reduction of approximately 2.5 points compared to the original texts. This decrease indicates a significant enhancement in readability, making the texts more accessible to users with varying levels of expertise.

The SARI scores, which measure content retention and simplification effectiveness, ranged from 39.76 to 40.58. These results suggest that our approach effectively preserved key information while simplifying the text. High SARI scores indicate that the essential content of the original texts was maintained, even with significant simplifications.

BLEU scores, which assess the alignment between the rewritten texts and reference versions, varied from 5.46 to 7.53. While these scores are modest, they demonstrate that our rewritten texts align reasonably well with reference texts, indicating effective rewriting while focusing on readability.

Compression ratios, which measure the proportion of text reduction, ranged from 0.90 to 1.17. These ratios show that our method effectively managed text length, avoiding excessive compression while maintaining readability. Sentence splits averaged around 1.37 per text, reflecting that our method preserved sentence structure while simplifying content.

Levenshtein similarity scores, ranging from 0.51 to 0.56, suggest that our method made meaningful changes to the original texts while retaining essential information. The moderate similarity scores align with our goal of rewriting rather than simply rephrasing.

Proportions of additions and deletions were 0.48 and 0.58, respectively, indicating a balanced approach to modifying the original text. Lexical complexity scores ranged from 8.34 to 8.51, suggesting that some degree of complexity was retained despite simplifications.

When compared with other methods, our approach showed competitive performance. For instance, the baseline method (References) achieved an FKGL score of 8.91 and a SARI score of 100.00. Although our FKGL scores were slightly higher, our SARI scores were close, indicating effective content preservation.

Elsevier's methods, such as Elsevier@SimpleText_Task3.1_run1, reported FKGL scores of 10.33 and SARI scores of 43.63. While our FKGL scores were lower, demonstrating better readability, their higher SARI scores suggest slightly superior content preservation. Similarly, UAms methods, like UAms_Task3-1_GPT2_Check, had FKGL scores of 11.47 and SARI scores of 29.91. Our method outperformed this approach in readability, as indicated by lower FKGL scores.

The results of Task 3 validate the effectiveness of the LLAMA framework in simplifying scientific texts while preserving critical content. The substantial reduction in FKGL scores and competitive SARI scores reflect our success in enhancing readability and content retention. Future work will focus on improving alignment with reference texts and optimizing the balance between readability and information preservation to further refine our approach.

4. Discussion and Conclusions

This study, situated within the SimpleText track of CLEF 2024, aimed to enhance the accessibility of scientific texts through three core tasks: selecting relevant passages for simplification, identifying and explaining difficult concepts, and rewriting complex sentences. Our findings across these tasks reveal both the effectiveness of our methods and areas where further refinement is needed.

In Task 1, we focused on selecting relevant passages from scientific documents to include in simplified summaries. By leveraging ElasticSearch for querying and relevance scoring, we effectively identified passages that aligned well with user queries. However, while our approach demonstrated strong performance in retrieving relevant content, the precision of relevance scoring could be improved. Future work should aim to refine the relevance assessment process, potentially integrating more advanced models or hybrid approaches to enhance accuracy.

Task 2 involved identifying difficult scientific terms and providing explanations to aid comprehension. Utilizing named entity recognition and custom rule-based approaches, we successfully extracted and explained challenging concepts. The integration of Wikipedia definitions further enriched the explanations, but the results highlighted the need for more nuanced difficulty rating mechanisms. Moving forward, incorporating advanced semantic analysis and contextual understanding could improve the accuracy of difficulty ratings and the quality of explanations.

In Task 3, our primary objective was to simplify scientific texts while preserving essential information. The LLAMA framework proved effective in reducing text complexity, achieving substantial improvements in readability. Despite these advances, our approach faced challenges in maintaining perfect fidelity to the original content. The balance between simplification and preservation of critical details is crucial. Future refinements should focus on enhancing the alignment with reference texts and ensuring that simplifications do not compromise the accuracy of the information.

When comparing our methods to others in the field, our approach exhibited competitive performance across all tasks. While some methods achieved slightly better results in specific areas, our overall approach demonstrated a robust balance between readability improvement and content preservation. The variations observed in different methods' outcomes underscore the complexity of text simplification and the need for a tailored approach depending on specific requirements and contexts.

Our research contributes valuable insights into the simplification of scientific texts. The results affirm the effectiveness of our methods in enhancing text accessibility, though they also reveal areas for improvement. The successful reduction in complexity across tasks highlights the potential for our approach to make scientific literature more approachable and engaging for a wider audience. To build on our findings, future research should focus on refining the precision of relevance scoring in passage selection, enhancing difficulty rating mechanisms in concept identification, and improving the balance between simplification and content preservation in text rewriting [7]. Incorporating user feedback and exploring advanced machine learning techniques will be essential for further advancing the accessibility of scientific texts. By addressing these areas, we can enhance the effectiveness of text simplification methods and contribute to a more inclusive dissemination of scientific knowledge.

Acknowledgments

We thank the CLEF 2024 SimpleText track organizers for their guidance and support. Special thanks to Liana Ermakova for her overview of modern methods and training students, including those with no coding background, as part of the BIP course AI for Humanitarians. We also thank the course organizers for providing the opportunity to exchange experiences and knowledge with students from various European universities.

References

[1]. Liana Ermakova et al. 2024. Overview of CLEF 2024 SimpleText Track on Improving Access to Scientific Texts. In Lorraine Goeuriot et al. (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024). LNCS. Springer-Verlag.

[2]. Liana Ermakova et al. 2024. Overview of the CLEF 2024 SimpleText Task 3: Simplify scientific text. In: Guglielmo Faggioli et al. (Eds). 2024. Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024). CEUR Workshop Proceedings, CEUR-WS.org.

[3]. Liana Ermakova, Eric SanJuan, S. Huet, H. Azarbonyad, Di Nunzio, G. M., F. Vezzani, , ... & J. Kamps, (2024). CLEF 2024 SimpleText Track: Improving Access to Scientific Texts for Everyone. In Advances in Information Retrieval: 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24–28, 2024, Proceedings, Part VI. Springer-Verlag.

[4]. Vinod S.S. Chandra, and Hareendran Anand S. Artificial Intelligence and Machine Learning. PHI Learning, 2014. ISBN 8120349342, 9788120349346.

[5]. Eric SanJuan et al. 2024. Overview of the CLEF 2024 SimpleText Task 1: Retrieve passages to include in a simplified summary. In: Guglielmo Faggioli et al. (Eds). 2024. Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024). CEUR Workshop Proceedings, CEUR-WS.org.

[6]. Giorgio Maria Di Nunzio et al. 2024. Overview of the CLEF 2024 SimpleText Task 2: Identify and explain difficult concepts. In: Guglielmo Faggioli et al. (Eds). 2024. Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024). CEUR Workshop Proceedings, CEUR-WS.org.

[7]. John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Illustrated edition. The MIT Press, 2015. ISBN 0262029448, 9780262029445.

[8]. Chris Sebastian. Python Programming for Beginners: Learn Python Machine Learning Language from Scratch, Deep Learning with Python. Amazon Digital Services LLC - KDP Print US, 2018. ISBN 1792874650, 9781792874659.

Appendix

In this section, we provide additional details on the prompts used for each task in our experiments. The prompts include the queries used to retrieve relevant passages, the requests for difficulty ratings of scientific terms, and the instructions for generating simplified explanations. Additionally, any larger tables or figures that accompany the manual data and generated results per task can be found here for reference.

Table 1
Prompts for Task 1 - Passage Selection

Query ID	Query Text
1	"Assigning female genders to digital assistants such as Apple's Siri and Amazon's Alexa is helping entrench harmful gender biases, according to a UN agency. Research released by Unesco claims that the often submissive and flirty responses offered by the systems to many queries – including outright abusive ones – reinforce ideas of women as subservient. Because the speech of most voice assistants is female, it sends a signal that women are obliging, docile and eager-to-please helpers, available at the touch of a button or with a blunt voice command like 'hey' or 'OK'," the report said. The assistant holds no power of agency beyond what the commander asks of it. It honours commands and responds to queries regardless of their tone or hostility. In many communities, this reinforces commonly held gender biases that women are subservient and tolerant of poor treatment."
2	…
3	…

Table 2
Prompts for Task 2 - Difficulty Identification

Term ID	Scientific Term	Prompt for Difficulty Rating
1	DNA replication	"Please rate the difficulty of understanding the term 'DNA replication'."
2	Quantum mechanics	"Rate the complexity level of the term 'Quantum mechanics'."
3	Electrochemical cell	"Provide a rating for the comprehensibility of 'Electrochemical cell'."

Table 3
Prompts for Task 3 - Text Simplification

Sentence ID	Original sentence	Prompt for simplification
1	"The mitochondria is the powerhouse of the cell."	"Simplify the following sentence: The mitochondria is the powerhouse of the cell."
2	"Photosynthesis is the process by which plants make food."	"Please simplify: Photosynthesis is the process by which plants make food."
3	"The theory of relativity revolutionized modern physics."	"Simplify this statement: The theory of relativity revolutionized modern physics."