# NLPalma @ CLEF 2023 SimpleText: BLOOMZ and BERT for Complexity and Simplification Task

Notebook for the SimpleText Lab at CLEF 2023

Victor Manuel Palma Preciado [1,2], Carolina Palma Preciado [1] and Grigori Sidorov [1]

[1] *Instituto Politécnico Nacional de México, Gustavo A. Madero, Ciudad de México, México*
[2] *Université de Bretagne Occidentale, HCTI, France*

**Abstract**

The following work has the purpose of describing the participation in the SimpleText 2023 track, on the identification of the term and its identification of the terms and their difficulty as a term, among themselves, all this belonging to Task 2. To solve this task, we used an approach of language models using Bloom but opted for its BLOOMZ version for a fine-tuning more focused on human instructions or in a more understandable way with more description-style prompts given by text input on a task. To solve the handling of the difficulty between terms a very simple classifier based on BERT-multilingual was used since this was developed as a binary classification and for the term vs. term evaluation a small algorithm was taken to accommodate the internal term classifications. On the other hand, we also participated in Task 3 in which the objective was to simplify passages extracted from abstracts, using the same approach as Task 2, BLOOMZ was used for the simplification of this text since different prompts were tested in case it was necessary to make several passes with those parts that yielded poor results or null results. Given that this was the first time participating in such tasks we can say that the results obtained were quite satisfactory even though we believe that they can be substantially improved with some other approach, which would have to be further reviewed.

**Keywords**

simplification, concept identification, term difficulty, term classification

## 1. Introduction

For this work the main objective we seek to fulfill the tasks as completely as possible, we believe that language models can present a good starting point to develop the tasks of this track, therefore BLOOMZ [2] was taken as a model to use because its qualities allow us to have input more consistent with human instruction, This means that you have to be very clear with the instructions given to the model and therefore the prompt to be used, it is also important to provide the model with as much context as possible to get the desired results. In this case, the simplification of the passage can alter the results depending on the previous context given, the same happening if the prompt is altered.

A very important part that had to be taken into account is the extraction of important terms in the given sentence for the task, besides assigning a number according to its complexity, so this complexity can have different parameters to measure the level of complexity whether it is the size of the word, the linguistic context, among others, so taking any of them could perhaps impact this type of classification based on complexity, so it was decided to opt for a binary classification.

[1]CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece
EMAIL: victorpapre@gmail.com (A. 1); c.palma.p0@gmail.com (A. 2); sidorov@cic.ipn.mx (A. 3)
ORCID: 0000-0001-8711-1106 (A. 1); 0000-0003-3253-4464 (A. 2); 0000-0003-3901-3522 (A. 3)

It is true that the starting point is BLOOMZ but we also have to use other types of models to do classification, although BLOOMZ is able to do this classification with the correct prompt, we believe that the use of other tools can speed up the process without having to rely entirely on BLOOMZ in which the more parameters you try to take, the longer it will take to give results, obviously this in many cases allows better results despite the delay that could represent. To classify we use a BERT-type model, in this case the multilingual one, to foresee any eventuality of the dataset.

## 2. Approach to the task

From the task proposed by SimpleText [1], Task 2 and Task 3 were carried out by implementing NLP models specifically deep learning models like transformers. In Task 2 identification and explanation of difficult concepts in the paragraph are done with the aim to define the meaning of key concepts for better conceptualization and understanding of a paragraph. On the other hand, Task 3 aims to create summarized text from scientific abstracts written in English, this concise version should maintain the original meaning in a simplified form.

The data [3] provided is divided into three dataset sizes (small, medium, and large), for this approach it was decided to use the medium dataset since is the middle ground of a sample of the text and sufficient data to observe the performance of the applied method. Text summarization is a difficult task that can be achieved from different approaches, in this case, a deep learning model known as BLOOMZ & mT0 is used to generate text summarization and determine difficult words in a paragraph. Since there are multiple versions of this model with different parameters, different versions were used when performing the tasks as the resources available were not sufficient to run some models.

The first step to complete Task 2, was to identify up to five of the most difficult terms in a sentence the first pass was done by using mt0-xl since this version could be loaded in a local environment without running out of memory, the first pass resolved around half of the dataset (2,320 sentences). After having several empty results with various prompts like “Give me up to five of the most difficult words of the next text:”, “Give me up to five of the most difficult words of the next sentence:”, “Give me the top five most difficult words of the next text:”, and “Suggest me at up five of the most difficult words on the next text:” it was decided to do the second pass and third pass with the Hugging Face inference API [7] that uses the BLOOMZ finetuned model with 176B parameters.

To explain the terms obtained in the first step, the same model is now implemented with prompts such as “Meaning of:”, “Definition of:”, “Give me the definition:”, “Give me the meaning:” all of this prompt were used in succession if one of them did not produce results, the next prompt was used until a favorable output was obtained. Since the task accepts both the definition of the word and an example of it, no further process was done.

The next step consists in scoring and ranking the terms, for the first part a multilingual BERT [6] was trained with the corpus provided, the score in the training text varies from 1 to 2 so the terms were assigned this range of scoring, 2 indicating a difficult term and 1 an easier one. The BERT model trained for the scoring was executed with the help of the Ktrain [5] wrapper that allows to load and perform the fine-tuning process.

Then Term ranking is done using the score and the term length, fist the words of the sentences are found and separated by the score, then a sub-ranking is designated from the length of the words, where the term with the most characters is the most difficult and the term with the fewest is the easiest term. If the terms are the same length they are arranged in order of appearance. Finally, when both lists for scores 1 and 2 are ordered, they are united to create the final ranking.

To elaborate Task 3 BLOOMZ is also implemented, this time instead of using the inference API a system for inference and fine-tuning called Petals [8] is used since the Hugging Face API, this platform joins computer sources from different servers to increase the computational power. The summarization prompts include “Summarize the text:” and “Summarize the sentence:”. As with previous processes, multiple runs were executed to obtain the best results.

## 3. Resources employed

Among the resources used to train and evaluate the models, Google's Colab environment was used, this platform allows Python programming and execution. It also allows an easier use of GPU, which served to perform the described tasks faster since this resource allows performing multiple simultaneous computations. The server used has the following specifications GPU NVIDIA-SMI 525.85.12, CUDA v12.0, and 25 of RAM.

As part of the resources used, it was decided to use a combination of Google Collaboratory with Petals for Task 2 in complexity spotting and Task 3 for simplification of scientific text, since we do not have the computational capacity to run models of parameter size 176B, which represent a larger amount of computation than we can handle, so the use of Petals becomes indispensable for those jobs that do not have enough computational capacity since it allows the benefits of crowd computing to make an inference with this type of large models in a relatively easy/semi-efficient way and in a way having the flexibility of an API and the power of PyTorch.

On the other hand, for the complexity ranking, we used BERT-multilingual [6], and for the internal complexity between terms of the same sentence we used a small algorithm that takes the length of the sentence and the value obtained from the previous classification to make 2 groups, if necessary, between those that have the value of 1 and 2 in complexity to generate the internal complexity ranking.

## 4. Results

As part of the results that we will present below, we will explain from a very general point of view the type of results obtained for each task and some of their particularities. We will explain some of the cases that we believe could be interesting to take into account for our future participation in these tasks for SimpleText in subsequent deliveries, taking into account that perhaps we could use some other more ad hoc models for tasks such as simplification.

### Task 2: “What is unclear?” Difficult concept identification and explanation

In the following example, we can see that the word that was extracted is CNN, but it did not know how to obtain the context of the sentence, so when throwing the meaning or the explanation it misunderstands CNN (Convolutional Neural Network) for CNN (Cable News Network) and this is completely incorrect, since as we can infer they are completely different things, one refers to a Machine learning model and the other to a news network, so there is no point of comparison and the result obtained refers to the latter.

- A conventional CNN for the end-to-end control is designed to map a single front-facing camera image to a steering command.
- Term rank: 2
- Word: CNN
- Word Meaning or explanation: CNN.com

It is logical to think that cnn.com refers to the web address of the news channel and clearly not to the type of neural network since there is no way that such a domain refers to an abstract element rather than a company.

The following example is one that we believe represents a term that could be difficult to explain, since being acronyms these can mean different things, which will depend on the field of study that makes reference to the context of the sentence, so it is extremely important to take this into account and in this case, it is clear that the training data off ease in the way of how to find a difficult term and therefore the context that has these acronyms, throw a correct explanation.

- This work explores the electromagnetic (EM) side-channel analysis literature for the purpose of assisting digital forensic investigations on IoT devices.

- Term Rank: 2
- Word: IoT
- Word Meaning or explanation: Internet of Things (IoT) is a network of physical objects or "things" embedded with electronics

It could also refer to another acronym such as EM but since this term could be complex, but is already explained in that sentence, it would not make much sense to take it into account.

The classification was another aspect that had to be applied in the sentences, this is a classification for the term extracted given its complexity and if more than one term was extracted from the sentence these had to be classified among themselves obtaining an ascending value given its complexity, this can be seen in the following words extracted from the same passage:

- attack Term difficulty: 2 Internal term difficulty: 1
- system Term difficulty: 2 Internal term difficulty: 2
- blind Term difficulty: 2 Internal term difficulty: 3
- False data Term difficulty: 1 Internal term difficulty: 4
- injection Term difficulty: 1 Internal term difficulty: 5

It is observed that the length of the word does not play such an important role in the rank assigned to a term under its difficulty, the explanation can only be inferred but that certain words are ranked higher than others is due to the role they play in the context and the difficulty to explain them in the given context.

## Task 3: Rewrite this! Rewriting scientific text

The case of simplifying a passage is a very interesting one, the idea of compacting a sentence or idea to such an extent, can be complicated, since this can lead, if it is very compact, to lose information and not be able to fully understand the idea or lend itself to an erroneous interpretation, on the other hand, that the idea is quite loose and does not explain things well, by summarizing incorrect ideas.

As can be seen, the following sentence was summarized erroneously since, as can be seen, it loses the focus of the information and does not mention that the data were collected by researchers, the center of information is the process, not the people who collected it.

- Excerpt: Permissions were taken from required authorities who made sure that minimal risk (to pedestrians) is involved in the data collection process.
- Simplified excerpt: The data was collected by a team of trained researchers

In this positive case of simplification, we can observe the type of simplification of the information in terms of content while maintaining the context and even explaining it in much simpler terms, instead of saying that the information is produced and accessed, it changes it to a source of information, which in this context means the same thing, although the final part of the sentence could be omitted and would not affect the result producing a more compact fragment.

- Excerpt: The advent of WWW changed the way we can produce and access information.
- Simplified excerpt: The Web has become a major source of information for many people

## 5. Conclusions

We can conclude that our decisions in the approach taken can give us good results and observe that they can be improved in a substantial way in which the results are presented, for example in the simplification task it can be observed that many times the context of the original sentence is lost, perhaps giving more information about the same sentence could solve this problem, On the other hand, for the task of term extraction, we could observe that it is difficult to obtain an accurate way of knowing that

the term is really difficult in terms of its context or that perhaps it is a compound term or an acronym that makes it complex, this leaves open different possibilities for improvement that we believe can be exploited in subsequent works.

## 6. References


[1] Liana Ermakova, Eric SanJuan, Stéphane Huet, Olivier Augereau, Hosein Azarbonyad, 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] Muennighoff, N., Wang, T., Sutawika, L., Roberts, A., Biderman, S., Scao, T. L.,et.al. (2022). Cross-lingual generalization through multitask finetuning. ArXiv Preprint ArXiv:2211. 01786.S. Cohen, W. Nutt, Y. Sagic, Deciding equivalances among conjunctive aggregate queries, J. ACM 54 (2007). doi:10.1145/1219092.1219093.

[3] Ermakova, Liana, Eric SanJuan, Stephane Huet, Olivier Augereau, Hosein Azarbonyad, and Jaap Kamps. "CLEF 2023 SimpleText Track: What Happens If General Users Search Scientific Texts?" ECIR 2023 Proceedings.

[4] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR, abs/1810.04805. Retrieved from http://arxiv.org/abs/1810.04805I.

[5] Arun S. Maiya (2020). ktrain: A Low-Code Library for Augmented Machine Learning. arXiv preprint arXiv:2004.10703.

[6] Pires, T., Schlinger, E., & Garrette, D. (2019). How Multilingual is Multilingual BERT? https://doi.org/10.18653/v1/p19-1493

[7] Inference API - Hugging Face. (s. f.). https://huggingface.co/inference-apiCañete, J., Chaperon, G., Fuentes, R., Ho, J.H., Kang, H., & Pérez, J. (2020). Spanish PreTrained BERT Model and Evaluation Data. In PML4DC at ICLR 2020.

[8] Borzunov, A., Baranchuk, D., Dettmers, T., Ryabinin, M., Belkada, Y., Chumachenko, A., Samygin, P., & Raffel, C. (2022). Petals: Collaborative Inference and Fine-tuning of Large Models. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2209.01188