Automatic Simplification of Drug Descriptions

Alexandr Lagornîi, Victoria Bobicev, Victoria Alexei
Technical University of Moldova, alexandr.lagornii@iis.utm.md, victoria.bobicev@ia.utm.md, victoria.alexei@iis.utm.md, ORCID: 0000-0003-4450-3964, 0000-0003-4560-3131

Keywords: automatic text simplification, drug description

Abstract. Automatic simplification of drug descriptions and medical texts is a very important task, it helps people with low medical literacy to understand complex drug descriptions and medical texts. Automatic text simplification is the process of converting source text into another text that conveys the information of the source text but is easier to understand and read [3]. This process usually involves replacing complex or unknown phrases or words with simple equivalents and converting long and syntactically complex sentences into shorter and less complex ones [1].

Automatic text simplification usually focuses on two methods of simplification, lexical and syntactic. Lexical text simplification seeks to replace words so that they are easier to understand, or to add appropriate definitions to a sentence. For example, “The book was great” would become “The book was excellent” or “The boy had tuberculosis” would become “The boy had tuberculosis, a lung disease.” Changing words in context is not a trivial task because the original meaning of the text can very easily be lost or misinterpreted [2].

Syntactic simplification aims at identifying syntactic phenomena in sentences that may make readability and comprehension difficult, with the aim of possibly transforming the found sentences into more readable and understandable equivalents. For example: “The festival was held in New Orleans, which was recovering from Hurricane Katrina” is converted in two simpler sentences: “The festival was held in New Orleans. New Orleans was recovering from Hurricane Katrina”.

Experiments of this work fell into three main categories:

1. Experiments without the use of large language models used data such as word frequency and dictionaries to find compound words and their definitions.

2. Experiments with exclusive application of large language models to simplify a given drug description [3].

3. Combined experiments that solved certain problems with large language models while others without them. For example, we defined a word difficult to understand using a large language model and then inserted its definition from a dictionary.

In conclusion, both methods are useful for the simplification: the wordfreq library identifies complex words as well as ChatGPT3.5 but ChatGPT is better at simplifying text due to good corpora training, while dictionary parsing is a very unreliable method that is also difficult to verify for correctness.

References

[1] Rémi Cardon, Natalia Grabar. 2020. French Biomedical Text Simplification: When Small and Precise Helps. In Proceedings of the 28th International Conference on Computational Linguistics, pages 710–716, Barcelona, Spain (Online). International Committee on Computational Linguistics.

[2] Sanja Stajner. Automatic Text Simplification for Social Good: Progress and Challenges. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 2637–2652, Online. Association for Computational Linguistics.

[3] Sergiu Nisioi, Sanja Stajner, Simone Paolo Ponzetto, Liviu P. Dinu. 2017. Exploring neural text simplification models. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 85–91, Vancouver, Canada. Association for Computational Linguistics.