University *of Ljubljana*
Faculty *of Computer and Information Science*

Automatic generation of Slovenian traffic news for RTV Slovenija

Marjan Stojchevski, Gal Šubic, and Aleksander Osvald

Abstract

Keywords
Keyword1, Keyword2, Keyword3 ...

Advisors: Slavko Žitnik

Introduction

The goal of this project is to automate the writing of Slovenian radio traffic reports from traffic information. These are currently being manually written by students and since this is a repetitive task that involves processing large amounts of data, we want to automate it using increasingly popular large language models (LLM's). Reports generated this ways should not only be factually correct and concise, but should also stick to established report form and naming conventions. Generating text in Slovene also poses a challenge, as most LLM's and their evaluation methods are better suited for English language.

Related works

G. Taghizadeh [1] talks about how how multi agent LLMs are affecting reporting, J. Pereira et al. adresses the broader field of news [2].

Dataset

The data for the project was already provided. It consists of three main parts: input data, output data (traffic reports) and rules and guidelines, which have to be taken into account when making reports. The input was collected from 2022 to 2024 from promet.si and the output consists of reports from RTV SLO.

The input data contains different categories of traffic information such as accidents, traffic jams, road work, weather related information and vehicle restrictions. The output data contains structured reports which are in accordance with rules and guidelines. They are further split into urgent reports (Nujna prometna informacija), which are broadcast when needed, and regular reports (Prometna informacija) which are broadcast at regular intervals every half hour.

The rules and guidelines are meant for human writers to correctly structure the report and use the proper terms and names. They include traffic information word structure, traffic event hierarchy, road and highway informal names, which are better understandable and more commonly used, and other relevant information.

Initial ideas

Although such a task could - to a degree - be accomplished using prompt engeneereng, it proves difficult to generate a report that accurately describes cause and effect, while sticking to a desired format.

At this stage, a basic preprocessing was done on both the input and output data. Html tags from the input data were stripped. The output RTF files were parsed and all output data was consolidated in a single file.

The first step is to find the relationship between the input and the output and which input generates which output. In the overall dataset there are more input samples than output samples and there are duplicates in the input samples. The date and time will be extracted from the output samples.

After the dataset is fully understood and preprocessed, the data will be prepared for feeding into a Transformer based Trainer by selecting a test, validation and train subsets.

At this stage, a model will be selected. A base model performing well on text-to-text generation tasks like Mistral or Llama2 will be tested with 4-bit quantization. A model pretrained on Slovenian language like "SambaLingo-Slovenian-Base" will also be evaluated.

While human evaluation would assess fluency and cor-

rectness of generated reports best, it is not an option for this assignment. We could use BERTScore [3] as an automatic measure of similarity between our models output and RTV SLO reports. BERTScore is a language-independent evaluation metric that measures the similarity between generated and reference text using contextual embeddings from transformer models. Alternative measures include the likes of BLEU, ROUGE, METEOR, but these are not the most suitable for Slovene due to its flexible word order and rich word morphology. A possible alternative is evaluation using other LLMs like GPT-4-turbo (ChatGPT) which understands Slovene language to compare the human-written and our machine-generated report. This approach could be especially useful for determining wheter the generated reports are in accordance with the reporting and naming guidelines. Lastly we could use some naive methods, such as comparing certain keywords that would have to appear in the report, such as names of locations and highways. For the test dataset we can use the last 6 months (equating to roughly 20%) of our dataset for evaluation or we can sample the reports evenly through time.

Methods

Results

Discussion

Acknowledgments

References

[1] How multi agent llms are revolutinizing reporting. https://medium.com/@gelareh.taghizadeh_63525/how-multi-agent-llms-are-revolutinizing-reporting-5eec0113146a. Accessed: 2025-03-21.

[2] João Pereira, Wenderson Wanzeller, and António Miguel Rosado da Cruz. Generation of breaking news contents using large language models and search engine optimization. In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024)*.

[3] Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*, 2020.