Vietnamese Text Summarization with LangChain and ChatGPT

Nguyễn Xuân Đông
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam
student639@ictu.edu.vn

***Abstract*—This undergraduate thesis investigates Vietnamese text summarization using LangChain and ChatGPT. Automatic summarization can help readers identify the main points of long Vietnamese documents. The student builds a summarization application and compares its output with ViT5 and BARTpho on the VNDS dataset using ROUGE-1, ROUGE-2, and ROUGE-L. The thesis reports higher ROUGE-1 and ROUGE-2 for the LangChain and ChatGPT approach than for the two compared models in its tests, while ViT5 and BARTpho performed better on ROUGE-L. The report does not establish performance on longer documents, broader domains, or independently repeated experiments. Its conclusion notes that processing long texts needs optimization. The contribution is a documented student study or prototype whose stated scope can guide further testing.**

***Index Terms*—Vietnamese summarization, LangChain, ChatGPT, ROUGE**

I. Introduction

Automatic summarization can help readers identify the main points of long Vietnamese documents. This thesis investigates Vietnamese text summarization using LangChain and ChatGPT. It begins from a practical engineering need and explains the technologies and workflow chosen for the case. The objective is to document a student implementation or comparative investigation, together with the evidence available in the manuscript. Broader operational or economic effects require separate measurements and are not assumed from the existence of a prototype [1]. Earlier ICTU student work on related designs is included here for context [2]. Related published studies provide context for the subject and methods [3], [4], [5], [6], [7], [8], [9], [10]. These contextual works were selected for this English summary; they are not asserted to appear in the original thesis bibliography.

II. Methods

The student builds a summarization application and compares its output with ViT5 and BARTpho on the VNDS dataset using ROUGE-1, ROUGE-2, and ROUGE-L. The source is summarized through its stated design, implementation material, examples, and conclusion. Features shown in the manuscript are distinguished from proposals for later development. This review does not add a new experiment, fill missing measurements with estimates, or treat a planned capability as an observed result. The source thesis remains the provenance for the reported work [1].

Supervisor: Trần Văn Khánh, Thai Nguyen University of Information and Communication Technology (tvkhanh@ictu.edu.vn).

III. Results

The thesis reports higher ROUGE-1 and ROUGE-2 for the LangChain and ChatGPT approach than for the two compared models in its tests, while ViT5 and BARTpho performed better on ROUGE-L. The report's screens, configurations, examples, or measurements constitute the evidence for this summary. A capability described only as a design intention is not counted as a validated operational outcome. No additional performance, adoption, financial, or safety figure has been introduced here. The result is interpreted within the scope of the student's documented setup [1].

IV. Discussion

The report does not establish performance on longer documents, broader domains, or independently repeated experiments. Its conclusion notes that processing long texts needs optimization. Further work could publish the exact test split and settings, evaluate factual consistency with human reviewers, and measure runtime on longer documents. These limitations affect how the evidence should be read. The current manuscript can support understanding of the design and reported demonstration, while independent tests would be needed for stronger claims. Any later deployment should document its environment and separate observed outcomes from intended benefits [1].

V. Conclusion

The thesis documents Vietnamese text summarization using LangChain and ChatGPT. The thesis reports higher ROUGE-1 and ROUGE-2 for the LangChain and ChatGPT approach than for the two compared models in its tests, while ViT5 and BARTpho performed better on ROUGE-L. The report does not establish performance on longer documents, broader domains, or independently repeated experiments. Its conclusion notes that processing long texts needs optimization. Further work could publish the exact test split and settings, evaluate factual consistency with human reviewers, and measure runtime on longer documents. Accordingly, the present contribution is a source-grounded account of the student work and its tested or illustrated scope. Claims about broader effectiveness should be assessed in later work using explicit evaluation conditions [1].

References

[1] Nguyễn Xuân Đông, ”Vietnamese Text Summarization with LangChain and ChatGPT,” internal undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2024.

[2] ICTU Student 197, ”Application of Genetic Algorithm in Text Classification of Vietnamese Language,” internal undergraduate thesis, Thai Nguyen University of Information and Communication Technology, 2010.

[3] Ti-Hon Nguyen et al., ”Text Summarization on Large-scale Vietnamese Datasets,” Journal of information and communication convergence engineering, 2022. doi:10.56977/jicce.2022.20.4.309.

[4] Ha Nguyen et al., ”Improving Quality of Vietnamese Text Summarization Based on Sentence Compression,” International Journal of Advanced Computer Science and Applications, 2016. doi:10.14569/ijacsa.2016.070250.

[5] Duy Ngoc Nguyen et al., ”Text summarization based on sentiment classification of comments from online Vietnamese newspaper,” Science and Technology Development Journal, 2016. doi:10.32508/stdj.v19i3.565.

[6] Sherry ., ”MULTILINGUAL TEXT SUMMARIZATION TECHNIQUES,” International Journal of Research in Engineering and Technology, 2017. doi:10.15623/ijret.2017.0607005.

[7] M. Kutlu et al., ”Generic Text Summarization for Turkish,” The Computer Journal, 2010. doi:10.1093/comjnl/bxp124.

[8] Swapnil Kuyate et al., ”AI Text Summarization System,” International Journal for Research in Applied Science and Engineering Technology, 2023. doi:10.22214/ijraset.2023.51481.

[9] Aarti Patil et al., ”Automatic Text Summarization,” International Journal of Computer Applications, 2015. doi:10.5120/19418-0910.

[10] Salvador Climent Roca, ”Automatic Text Summarization,” Digithum, 2001. doi:10.7238/d.v0i3.598.