

Research on the problem of named entity recognition (NER) in the legal domain

Ngo Thanh Kien

Abstract: The topic focuses on building a deep Named Entity Recognition (NER) model for Vietnamese legal texts using PhoBERT. The study trained the model to extract important information such as law names, articles, issuing authorities, and effective dates. The application of PhoBERT effectively handles the characteristics of the Vietnamese language and complex legal contexts. The achieved results serve as an important foundation for developing virtual legal assistant systems and intelligent document retrieval.

Keywords: Named Entity Recognition (NER), Legal Text, PhoBERT

References

- [1] D. Q. Nguyen and A. T. Nguyen, “PhoBERT: Pre-trained language models for Vietnamese,” Findings of EMNLP, pp. 1037–1042, 2020. [2] H. P. Nguyen, T. H. Nguyen, and T. H. Nguyen, “PhoNLP: A joint multi-task learning model for Vietnamese NLP,” Proc. EMNLP, 2021. [3] N. T. Huyền and P. A. Tuấn, “Trích xuất thực thể và quan hệ trong văn bản tiếng Việt,” Kỹ yếu Hội nghị FAIR, 2023.