

Transforming Indonesian Geography Education Books into Knowledge Graphs Using ChatGPT LLMs

1st Satria Siwi
School of Computing
Telkom University
Bandung, Indonesia

satriaajips@student.telkomuniversity.ac.id

2nd Kemas Wiharja*
School of Computing
Telkom University
Bandung, Indonesia

bagindokemas@telkomuniversity.ac.id

3rd Raihan Hafizh
School of Computing
Telkom University
Bandung, Indonesia

rhnatsal@student.telkomuniversity.ac.id

I. INTRODUCTION

Digital transformation in education is revolutionizing the dissemination and consumption of knowledge. Among various technological advancements, natural language processing (NLP) and artificial intelligence (AI) applications have shown significant potential to enhance learning experiences [1], [2]. This paper explores an innovative approach to converting Indonesian Geography Education Books, available in PDF format, into comprehensive Knowledge Graphs [3]. Utilizing the advanced capabilities of ChatGPT Large Language Models (LLMs) [4] and visualizing these graphs through Neo4j [5], we aim to create a more interactive and accessible learning tool.

Knowledge graphs work by structuring information into entities (such as locations, people, and events) and their relationships, allowing users to explore and understand complex data through visual and interactive means. This structured representation not only preserves the informational integrity of the original content but also enhances its accessibility and utility, making it particularly beneficial in educational contexts by facilitating easier navigation and comprehension of large datasets.

The dataset for this study consists of Indonesian Geography Education Books chosen for their rich and detailed content about Indonesia's geographical landscape, cultural heritage, and environmental features. These books, in PDF format, present a significant challenge due to their unstructured nature, making direct extraction of meaningful information difficult. To address this, we employ a meticulous process of text extraction, identifying and converting each word from the PDF documents into entities, relationships, and triplets, which form the foundation of a knowledge graph [6], [7].

Harnessing the power of ChatGPT LLMs, one of the most advanced models in natural language processing, this extraction process becomes highly efficient [8], [9]. ChatGPT's ability to understand and generate human-like text makes it exceptionally suitable for identifying entities such as locations, people, and events mentioned in the text, and for establishing relationships between these entities [10]. By leveraging ChatGPT's

sophisticated text processing capabilities, this research aims to extract and categorize entities accurately, establish meaningful relationships between them, and generate triplets that capture the semantic essence of the text.

Once the data is extracted, comprising entities and their interrelationships, it is used to construct a knowledge graph. Knowledge graphs provide a structured representation of information, showcasing the interconnections between different entities and offering a visual and interactive means to explore and understand complex information [11], [12], [13], [14]. For this purpose, Neo4j, a robust graph database management system, is employed to facilitate efficient storage, querying, and visualization of knowledge graphs [5]. To ensure the effectiveness and accuracy of the constructed knowledge graphs, we utilize two well-known evaluation metrics: Mean Reciprocal Rank (MRR) and HITS@K [15]. These metrics provide a comprehensive evaluation of both the ranking performance and the precision of the knowledge graph, ensuring that the constructed graphs are accurate and practical for educational purposes.

Demonstrating the feasibility and effectiveness of using ChatGPT LLMs for text extraction from PDF documents and the subsequent construction and visualization of knowledge graphs is the primary objective of this study. By transforming static, text-based educational materials into dynamic, interactive knowledge representations, this research aims to enhance the learning experience, making geographical knowledge more engaging and accessible to students and educators alike.

To validate the effectiveness of knowledge graphs in an educational setting, this study includes a case study conducted in a real classroom environment. The case study aims to provide empirical evidence of the impact of knowledge graphs on improving learning outcomes, showcasing how these tools can make complex information more accessible and engaging for students. By observing the practical application of knowledge graphs in a classroom, we aim to demonstrate their potential benefits in enhancing educational experiences. Additionally, we test the method using a wider range of data types and from different disciplines to assess its adaptability and effectiveness

* Corresponding author.

more broadly. By validating this approach across different text types and disciplines, we aim to demonstrate the generalizability of the method on a larger scale.

This paper is organized as follows, Section II provides a review of related work in the fields of knowledge graph construction and the application of NLP in education. Section III details the methodology employed in this study, including the processes of data extraction, entity recognition, relationship establishment, and knowledge graph construction. Section IV presents the results of the study, discusses the implications of the findings, and evaluates the knowledge graphs using MRR and HITS@K metrics. Finally, Section V concludes the paper and outlines potential directions for future research.

REFERENCES

- [1] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, and N. Zhang, "Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities," *arXiv preprint arXiv:2305.13168*, 2023.
- [2] Z. Bi, J. Chen, Y. Jiang, F. Xiong, W. Guo, H. Chen, and N. Zhang, "Codekgc: Code language model for generative knowledge graph construction," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 3, pp. 1–16, 2024.
- [3] X. Ma, "Knowledge graph construction and application in geosciences: A review," *Computers & Geosciences*, vol. 161, p. 105082, 2022.
- [4] S. Alaswad, T. Kalganova, and W. S. Awad, "Using chatgpt and other llms in professional environments," 2023.
- [5] H. Shatnawi and J. Saquer, "Encoding feature models in neo4j graph database," in *Proceedings of the 2024 ACM Southeast Conference*, 2024, pp. 157–166.
- [6] M. Rizun, "Knowledge graph application in education: A literature review," *Acta Universitatis Lodzianis. Folia Oeconomica*, vol. 3, no. 342, pp. 7–19, 2019.
- [7] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, "Knownedu: A system to construct knowledge graph for education," *IEEE Access*, vol. 6, pp. 31 553–31 563, 2018.
- [8] R. Omar, O. Mangukiya, P. Kalnis, and E. Mansour, "Chatgpt versus traditional question answering for knowledge graphs: Current status and future directions towards knowledge graph chatbots," *arXiv preprint arXiv:2302.06466*, 2023.
- [9] X. Q. Dao, N. B. Le, B. B. Ngo, and X. D. Phan, "Llms' capabilities at the high school level in chemistry: Cases of chatgpt and microsoft bing ai chat," 2023.
- [10] J. Zhang, "Graph-toolformer: To empower llms with graph reasoning ability via prompt augmented by chatgpt," *arXiv preprint arXiv:2304.11116*, 2023.
- [11] H. Paulheim, "Knowledge graph refinement: A survey of approaches and evaluation methods," *Semantic Web*, vol. 8, no. 3, pp. 489–508, 2017.
- [12] X. Zou, "A survey on application of knowledge graph," in *Journal of Physics: Conference Series*, vol. 1487, no. 1. IOP Publishing, 2020, p. 012016.
- [13] M. Kejriwal, "Knowledge graphs: A practical review of the research landscape," *Information*, vol. 13, no. 4, p. 161, 2022.
- [14] K. R. S. Wiharja, D. T. Murdiansyah, M. Z. Romdlony, T. Ramdhani, and M. R. Gandidi, "A questions answering system on hadith knowledge graph," *Journal of ICT Research & Applications*, vol. 16, no. 2, 2022.
- [15] C. T. Hoyt, M. Berrendorf, M. Galkin, V. Tresp, and B. M. Gyori, "A unified framework for rank-based evaluation metrics for link prediction in knowledge graphs," *arXiv preprint arXiv:2203.07544*, 2022.