Supplementary Materials of: ZeroEA: A Zero-Training Entity Alignment Framework via Pre-Trained Language Model

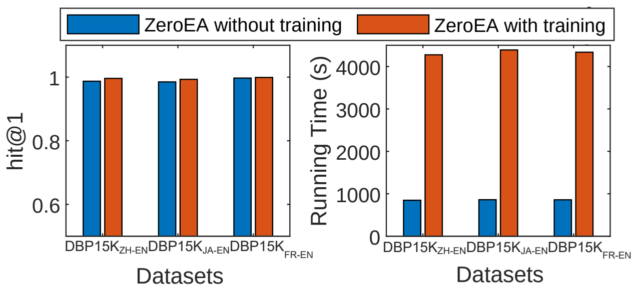

Figure 1: The performance and running time comparison over training and no training.

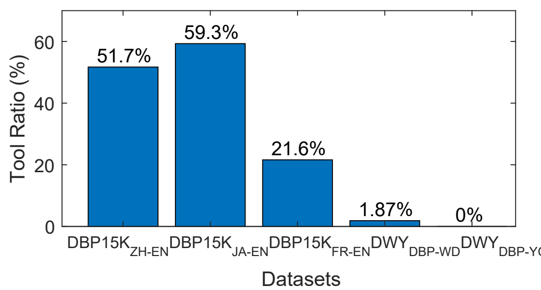

Figure 2: The percentage of WebSearch tool used entities in each dataset to achieve best performance.

1. Supplementary materials for the training version of ZeroEA.

For the training version, we regard the seeds (i.e., aligned entity pairs) as the positive examples and randomly choose an equal number of entity pairs as negative samples. We treat it as a binary classification task and utilize binary cross-entropy loss [1] to train. We divide the labeled pairs into training, validation, and test sets with a ratio of 6:1:3. We set the learning rate to 1e-5 and the batch size to 32.

Performance: After training, we observed an improvement in ZeroEA's performance, as depicted in the left part of Figure 1. However, it suffers from longer training time as shown in the right part of Figure 1. Our previous experiments have demonstrated that the prompts generated by our approach are highly effective in evoking knowledge in PLMs. The improved training results demonstrate that effectiveness is retained when training ZeroEA, which does not compromise the ability of our prompts. Furthermore, the growth observed with the trained ZeroEA shows the potential of our method to enhance training approaches, particularly in more complex real-world datasets.

2. Clarification on Biased Benchmarking & Infromation Leakage

We admit that it may leave the false impression that tool calls can only achieve better performance by revealing the ground truth answer. During conducting a thorough analysis and experiments, we discover that this issue is rare in our datasets (less than 1%). Also, we develop a strategy to avoid this information leakage when calling tools. The following sections will provide a detailed explanation of the tool call's motivation, the information leakage concern, its frequency in ZeroEA experiments, our strategy to prevent it, and the updated results that prove that our performance doesn't come from information leakage.

Motivation of Tool Calls. Tool calls leverage external resources to augment the reasoning capabilities of Pre-trained Language Models on complex tasks. They have proven beneficial across various domains and tasks, enhancing the capabilities of PLMs across multiple domains [4, 2, 3]. Tool calls also improve performance on EA tasks - for example, prior methods like RDGCN [6] and BERT-INT [5] used Google Translation on cross-lingual datasets like DBP15K. The rationale for using tool calls in our method is multi-faceted but does not include ground-truth answers, which could lead to biased settings. Primarily, we aim to supplement information about entities for PLMs to minimize entity recognition ambiguity, especially for long-tail (nodes whose degrees of a specific entity are low) cases. Also, our designed "Percipient" in Section 3.4 of the paper empowers ZeroEA to automatically determine when and where to use tools. This implies that tool calls don't require human intervention or application to all examples

Information Leakage. Using tools may leak the correct answer in the content retrieved from tools. This could lead to a misunderstanding among readers that the success of the ZeroEA framework is primarily due to tricks, rather than its real effectiveness. To resolve this concern, we carefully analyzed information leakage in all datasets by two procedures.1) We split entity name strings and response strings from tool calls into individual word strings. 2) Then we consider it as a leakage example if any of the entity name word strings appeared in the response word strings. The following are the comprehensive statistics related to data leakage in each dataset.

We find that the information leakage rate are:

1) $\mathrm{DBP15K_{ZH-EN}}$: **0.58%**;
2) $\mathrm{DBP15K_{JA-EN}}$: **0.79%**;
3) $\mathrm{DBP15K_{FR-EN}}$: **0.62%**;
4) $\mathrm{DWY100K_{DBP-WD}}$: **0.18%**;
5) $\mathrm{DWY100K_{DBP-YG}}$: **0.00%**.

Based on the above investigation, it can be observed that the instances of information leakage are remarkably rare across all the

datasets utilized in our study, with the highest rate not exceeding 1%. which indicates that experimental settings are unlikely to be biased by minor information leakage. To further mitigate potential negative impacts, we devised a strategy to remove contents from tool calls that could leak the correct answer, ensuring that the information introduced only supplements the target entity's information. After implementing this strategy, our model's performance only dropped slightly on two out of five datasets, remained unchanged on two, and even increased on one. The updated performance details after blocking leakage messages from tools:
1) $\mathrm{DBP15K_{ZH-EN}}$: 98.7%→98.5% (**↓0.2%**)
2) $\mathrm{DBP15K_{JA-EN}}$: 98.5%→98.2% (**↓0.3%**)
3) $\mathrm{DBP15K_{FR-EN}}$: 99.7%→99.8% (**↑0.1%**)
4) $\mathrm{DWY100K_{DBP-WD}}$: 99.8%→99.8% (-)
5) $\mathrm{DWY100K_{DBP-YG}}$: 99.9%→99.9% (-)

1. REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[2] L. Gao, A. Madaan, S. Zhou, U. Alon, P. Liu, Y. Yang, J. Callan, and G. Neubig. PAL: program-aided language models. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202, pages 10764–10799. PMLR, 2023.

[3] M. Komeili, K. Shuster, and J. Weston. Internet-augmented dialogue generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 8460–8478. Association for Computational Linguistics, 2022.

[4] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, L. Zettlemoyer, N. Cancedda, and T. Scialom. Toolformer: Language models can teach themselves to use tools. *arXiv preprint arXiv:2302.04761*, 2023.

[5] X. Tang, J. Zhang, B. Chen, Y. Yang, H. Chen, and C. Li. Bert-int: a bert-based interaction model for knowledge graph alignment. *interactions*, 100:e1, 2020.

[6] Y. Wu, X. Liu, Y. Feng, Z. Wang, R. Yan, and D. Zhao. Relation-aware entity alignment for heterogeneous knowledge graphs. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence, 2019.