RefineEA: A Modular Framework for LLM-Guided Refinement in Entity Alignment

Rumana Ferdous Munne[1] Md Mostafizur Rahman[2]
Yuji Matsumoto[1]
[1]RIKEN Center for Advanced Intelligence Project, Tokyo, Japan
[2]Rakuten Institute of Technology, Rakuten Group, Inc., Tokyo, Japan
{rumanaferdous.munne,yuji.matsumoto}@riken.jp, mdmostafizu.a.rahman@rakuten.com

Abstract

Entity alignment (EA) aims to identify equivalent entities across heterogeneous knowledge graphs (KGs). While existing embedding and graph based methods are scalable, their performance is often limited by data sparsity, structural heterogeneity, and semantic ambiguity. Large language models (LLMs) offer strong reasoning capabilities over unstructured and semi-structured information and have the potential to address these challenges; however, their application to entity alignment remains largely unexplored and is constrained by high computational cost, making end-to-end alignment impractical. To address this, we introduce RefineEA, a modular framework that employs LLMs as semantic judges to refine and improve the predictions of existing entity alignment models.

1 Introduction

Entity alignment refers to the task of identifying and linking equivalent entities across different knowledge graphs. Knowledge graph provides structured representations of real-world entities; however, differences in ontologies, naming conventions, attribute completeness, and relational structure make EA a challenging problem. Existing EA methods typically rely on multi-aspect entity embeddings. While effective in certain scenarios, these methods often struggle in the presence of sparse information, structural heterogeneity, and subtle semantic ambiguity.

Recent advances in large language models (LLMs) have introduced new opportunities for entity alignment. Trained on vast corpora of textual and factual knowledge, LLMs possess strong semantic understanding and reasoning capabilities that can complement traditional alignment signals. Several studies have explored incorporating large language models into EA pipelines, such as using LLMs for iterative reasoning over entity descriptions or for validating candidate alignments generated by embedding-based models [1]. These works demonstrate the promise of LLMs in improving alignment quality, particularly in complex or low-resource settings. However, directly applying LLMs to end-to-end entity alignment remains computationally expensive and difficult to scale.

To address these challenges, we propose RefineEA, a modular framework for LLM-guided refinement in entity alignment. Rather than replacing existing EA systems, RefineEA uses LLMs as semantic judges to refine and validate predictions from a wide range of EA approaches, including embedding-based, graph-based, and hybrid models—thereby improving accuracy while maintaining scalability.

Given candidate alignments produced by an underlying EA model, RefineEA constructs a compact contextual profile for each entity using its name, attributes, relational context, and textual descriptions. For each candidate alignment, the LLM reasons over the paired profiles to assess semantic consistency and estimate alignment reliability. A confidence-based decision strategy then integrates the LLM's assessment with the original model prediction: candidates that the LLM confidently identifies as incorrect are rejected, while uncertain cases defer to the original alignment. This design enhances robustness and balances precision and recall without over-reliance on LLM outputs. This approach ensures that only the most reliable alignments are selected, while maintaining the alignment process's accuracy. The key contributions are -

- We propose RefineEA, a modular framework that uses LLMs to refine entity alignment predictions without replacing existing models.

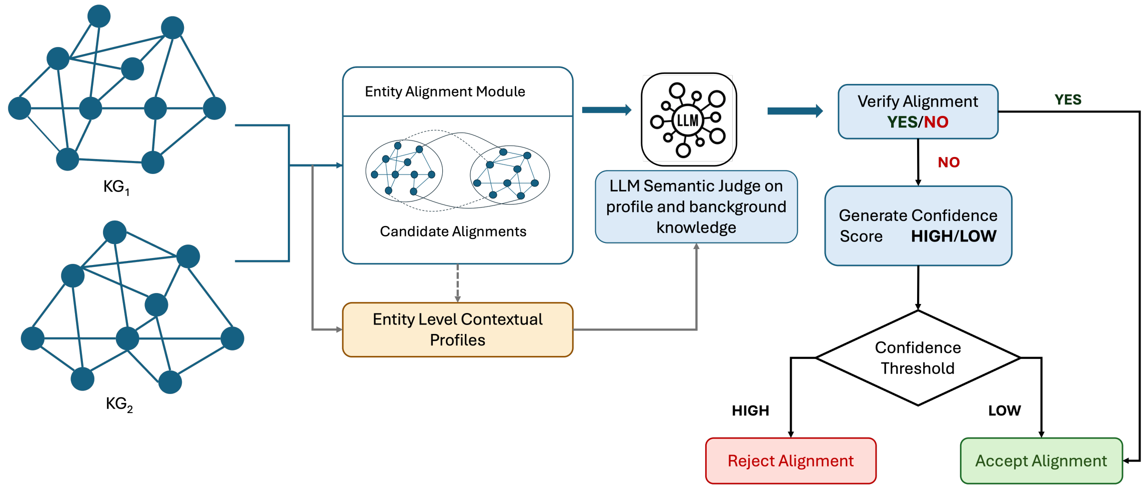

Figure 1 RefineEA Framework

- We introduce entity-level contextual profiles to support LLM-based semantic reasoning.
- We propose a confidence-based decision mechanism to resolve conflicts between the LLMs the base model's prediction, effectively handling uncertain cases.
- RefineEA is method-independent, making it applicable across diverse EA approaches and improving overall robustness.

Overall, RefineEA provides an efficient and scalable way to incorporate LLM reasoning into entity alignment, improving accuracy in challenging and heterogeneous settings.

2 Related Works

Entity alignment (EA) has been widely studied using embedding-based and graph-based approaches. Embedding-based methods represent entities and relations as low-dimensional vectors in a shared space, modeling relations as translations between entity embeddings [4]. Models like MTransE [2], BootEA [13], JAPE [12], KECG [8], TransEdge [14], JarKA [3], embed entities and relations into a shared vector space to map entities across different knowledge graphs.

GNN-based models like GCN-Align [18] and, RDGCN [5] leverage GNNs to integrate structural similarity and local graph neighborhoods, achieving comparable results. Yang et al. [20, 22, 21] also explore GNN-based and unsupervised approaches for entity alignment. BERT-INT [15] , JSAE [10] and EASAE [11] employ pre-trained language models to capture semantic richness from entity descriptions, while MultiKE [23], TEA [24] and SDEA [25] utilize multi-view information for alignment. Recent studies, such as Simple-HHEA [7] uses structured embeddings with logical reasoning.

ChatEA [26] integrates LLMs for iterative reasoning to filter candidates, leveraging names, descriptions, and structures and LLMEA [19] integrates multiple similarity features with LLM-based refinement. In contrast, our proposed RefineEA improves alignment accuracy, robustness, and interpretability by serving as a modular post-alignment refinement layer. It leverages LLM-guided reasoning over contextual entity profiles to handle data sparsity, semantic ambiguity, and heterogeneity, while remaining applicable across diverse EA models without changing their architectures.

3 Method

RefineEA is a modular framework designed to refine candidate entity alignments produced by existing EA models using large language model (LLM) reasoning. Instead of performing end-to-end alignment, RefineEA operates as a post-alignment refinement layer that that takes candidate alignments produced by an existing EA model and validates, corrects, or defers alignment decisions based on semantic evidence and confidence estimates. Figure 1 illustrates the overall architecture. RefineEA consists of three components: (1) contextual profile construction, (2) LLM-enhanced alignment verification, and (3) confidence-based decision integration.

3.1 Contextual Profile Construction

Knowledge graphs often exhibit heterogeneous representations, incomplete or inconsistent attributes, ambiguous descriptions, and even cross-lingual differences, which make entity alignment difficult—especially when semantic evidence is sparse. To support semantic comparison under these conditions, RefineEA converts each entity into a compact, text-based contextual profile that integrates key KG evidence into a unified, human-readable form. The profile includes the entity name (and aliases when available), important attributes, nearby relations, and any available descriptions. By turning scattered KG signals into a single, consistent representation, RefineEA enables more reliable semantic comparison even when the two KGs differ in schema or completeness.

3.2 LLM-enhanced Alignment Verification

Given a candidate entity pair (e_1, e_2) produced by a base entity alignment model, RefineEA uses an LLM to verify whether the two entities refer to the same real-world entity. Concretely, RefineEA feeds the paired entity profiles into the LLM through a structured verification prompt, asking for a binary decision based on semantic consistency between the profiles and the LLM's background knowledge.

Prompt 1 : Alignment Verification

> Given two entity profiles from different knowledge graphs, determine whether the two entities refer to the same real-world entity using background knowledge and contextual reasoning.
>
> **Respond with either:**
>
> - `[YES]` – if the profiles represent the same real-world entity.
> - `[NO]` – if the profiles represent different entities.

3.3 Confidence-based Decision Integration

If the LLM responds with `[YES]` in Prompt 1, RefineEA accepts the candidate pair as aligned. If the response is `[NO]`, RefineEA performs an additional confidence check to distinguish clear mismatches from uncertain cases. Specifically, RefineEA asks the LLM to provide a misalignment confidence score $s \in [0, 1]$, where a higher score indicates stronger evidence that the two entities are different. RefineEA then applies a confidence threshold CT (selected on a validation set) to integrate the LLM output with the base model prediction: when $s \geq CT$, the candidate alignment is rejected; otherwise, the LLM is treated as uncertain and RefineEA defers to the base model's original prediction.

Prompt 2: Misalignment Confidence (Only if response is `[NO]`)

> If you responded with `[NO]` in Prompt 1, provide a confidence score between `0` and `1` indicating how strongly the two entities are **not** aligned.
>
> **Scoring guide:**
>
> - A score close to `1` indicates strong evidence that the entities are different.
> - A score close to `0` indicates uncertainty or partial overlap.
>
> **Output format:** `[SCORE]={0..1}`

4 Experiment

In this section, we describe the datasets, model configurations, evaluation metrics, and experimental results.

4.1 Dataset and Implementation

Table 1 Dataset Statistics.

Dataset		#Entities	#Relations	#Anchor
ICEWS-WIKI	ICEWS	11,047	272	5,058
	WIKI	15,896	226	
ICEWS-YAGO	ICEWS	26,863	272	18,824
	YAGO	22,734	41	

We conducted experiments on two challenging entity alignment datasets [26], summarized in Table 1. ICEWS-WIKI and ICEWS-YAGO exhibit substantial heterogeneity between the paired KGs, including mismatched entity counts and differing graph structures, making alignment significantly more difficult than on more homogeneous benchmarks. In contrast, datasets such as DBP15K or DB-WIKI are often considered relatively "easy" due to stronger overlap in entities and schema, and have been extensively studied; as a result, improvements on them can be less indicative of robustness under severe distribution shifts. By evaluating on ICEWS-WIKI and ICEWS-YAGO, we aim to demonstrate that our method remains effective when textual descriptions and structural cues are sparse, inconsistent, or misaligned, which is the setting most relevant to real-world EA. In our experimental setup,

Model	ICEWS-WIKI	ICEWS-YAGO
TEA-GNN	0.126	0.064
+ **RefineEA**	**0.176**	**0.089**
BERT	0.596	0.784
+ **RefineEA**	**0.775**	**0.906**
Simple-HHEA	0.754	0.870
+ **RefineEA**	**0.844**	**0.924**
FuAlign	0.361	0.423
+ **RefineEA**	**0.469**	**0.528**
BERT-INT	0.607	0.793
+ **RefineEA**	**0.807**	**0.936**

Table 2 Entity alignment performance (MRR) on ICEWS-WIKI and ICEWS-YAGO.

we employed GPT-4 for entity profile generation and reasoning. The dataset was divided into training and testing sets in a 3:7 ratio. The evaluation metrics used were Hits@k (with $k = 1$ and $k = 10$) and Mean Reciprocal Rank (MRR).

4.2 Result and Discussion

To evaluate RefineEA, we apply it as a post-alignment refinement layer on top of several strong entity alignment baselines using a two-stage evaluation setup. We first generate candidate alignments using TEA-GNN, BERT, FuAlign, BERT-INT, and Simple-HHEA. RefineEA then verifies each candidate alignment by prompting an LLM to compare the paired entities using their contextual profiles and the LLM's background knowledge. If the LLM confirms the alignment with `[YES]`, the candidate is accepted. If the LLM rejects it with `[NO]`, we examine the confidence score: candidates with high misalignment confidence are removed, while low-confidence cases are treated as uncertain and we keep the original prediction from the base EA model. As shown in Table 2, RefineEA consistently improves alignment quality across all baselines. TEA-GNN shows the largest gains, indicating robustness to structural sparsity; BERT and BERT-INT improve by resolving semantic ambiguities beyond text similarity; FuAlign improves noticeably, suggesting reduced sensitivity to embedding noise; and Simple-HHEA shows smaller but steady gains, indicating that residual semantic errors can still be corrected even in strong hybrid models. Overall, these results suggest that contextual profiles provide an additional semantic signal complementary to embedding- and graph-based matching, especially under sparsity, heterogeneity, or ambiguous entity representations. Importantly, RefineEA does not replace the base alignment modules; it refines its output, making it applicable across different EA families without requiring changes to their training objectives or architectures. While recent LLM-based EA approaches can achieve strong results by relying heavily on LLM reasoning, they are often designed as end-to-end or tightly integrated pipelines, which reduces portability across different EA systems and may increase computational cost. In contrast, RefineEA is explicitly built as a refinement module: it attaches to the output of an arbitrary EA model and improves robustness through confidence-aware correction, providing a practical way to incorporate LLM reasoning into entity alignment without restructuring the underlying EA models.

We analyzed the impact of varying the confidence threshold (CT) from 0.60 to 0.95 using the random validation samples. When the LLM assigns a misalignment confidence score below CT, we treat this case as uncertain and defer to the underlying EA model. Across datasets, setting between 0.85 and 0.90 yielded the best balance.

5 Conclusion

In this paper, we introduced RefineEA, a modular post-alignment refinement framework that enhances entity alignment by incorporating large language model reasoning without replacing existing alignment systems. RefineEA constructs compact contextual entity profiles and uses LLM-based semantic verification with a confidence-aware decision mechanism to refine candidate alignments produced by diverse EA models.

Experimental results on heterogeneous knowledge graphs demonstrate RefineEA consistently improves alignment performance across multiple strong baselines. It shows substantial relative gains while preserving scalability and efficiency. Unlike end-to-end LLM-based approaches, RefineEA operates as a lightweight and method-independent refinement layer, making it practical for integration for existing EA methods.

For future work, we plan to adapt RefineEA to domain-specific entity alignment (e.g., biomedical or scientific knowledge graphs) by adding domain-aware profiles and prompts, and by tuning the confidence threshold to better handle specialized terms and ontology differences.

References

[1] Munne, Rumana Ferdous and Rahman, Md Mostafizur and Matsumoto, Yuji. Entity profile generation and reasoning with LLMs for entity alignment, In Findings of the Association for Computational Linguistics: EMNLP, pages 20073–20086, Novemebr 2025.

[2] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In IJCAI.

[3] Bo Chen, Jing Zhang, Xiaobin Tang, Hong Chen, and Cuiping Li. 2020. Jarka: Modeling attribute interactions for cross-lingual knowledge alignment. In Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part I 24, pages 845–856. Springer.

[4] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. volume 26.

[5] Zhibin Chen, Yuting Wu, Yansong Feng, and Dongyan Zhao. 2022. Integrating manifold knowledge for global entity linking with heterogeneous graphs. Data Intelligence, 4(1):20–40.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL, pages 4171–4186.

[7] Xuhui Jiang, Chengjin Xu, Yinghan Shen, Yuanzhuo Wang, Fenglong Su, Zhichao Shi, Fei Sun, Zixuan Li, Jian Guo, and Huawei Shen. 2024. Toward practical entity alignment method design: Insights from new highly heterogeneous knowledge graph datasets. In Proceedings of the ACM on Web Conference 2024, pages 2325–2336.

[8] Chengjiang Li, Yixin Cao, Lei Hou, Jiaxin Shi, Juanzi Li, and Tat-Seng Chua. 2019. Semi-supervised entity alignment via joint knowledge embedding model and cross-graph model. Association for Computational Linguistics.

[9] Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. 2021. Boosting the speed of entity alignment 10×: Dual attention matching network with normalized hard sample mining. In Proceedings of the Web Conference 2021, pages 821–832.

[10] Rumana Ferdous Munne and Ryutaro Ichise. 2020. Joint entity summary and attribute embeddings for entity alignment between knowledge graphs. In International Conference on Hybrid Artificial Intelligence Systems, pages 107–119. Springer.

[11] Rumana Ferdous Munne and Ryutaro Ichise. 2023. Entity alignment via summary and attribute embeddings. Logic Journal of the IGPL, 31(2):314–324.

[12] Zequn Sun, Wei Hu, and Chengkai Li. 2017. Cross-lingual entity alignment via joint attribute-preserving embedding. In The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21–25, 2017, Proceedings, Part I 16, pages 628–644. Springer.

[13] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping entity alignment with knowledge graph embedding. In IJCAI, volume 18.

[14] Zequn Sun, Jiacheng Huang, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Transedge: Translating relation-contextualized embeddings for knowledge graphs. In The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18, pages 612–629. Springer.

[15] Xiaobin Tang, Jing Zhang, Bo Chen, Yang Yang, Hong Chen, and Cuiping Li. 2020. Bert-int: A bert-based interaction model for knowledge graph alignment. In IJCAI.

[16] Bayu Distiawan Trisedya, Jianzhong Qi, and Rui Zhang. 2019. Entity alignment between knowledge graphs using attribute embeddings. In Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 297–304.

[17] Chenxu Wang, Zhenhao Huang, Yue Wan, Junyu Wei, Junzhou Zhao, and Pinghui Wang. 2023. Fualign: Cross-lingual entity alignment via multi-view representation learning of fused knowledge graphs. Information Fusion, 89:41–52.

[18] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual knowledge graph alignment via graph convolutional networks. In Proceedings of the 2018 conference on empirical methods in natural language processing, pages 349–357.

[19] Linyao Yang, Hongyang Chen, Xiao Wang, Jing Yang, Fei-Yue Wang, and Han Liu. 2024. Two heads are better than one: Integrating knowledge from knowledge graphs and large language models for entity alignment. arXiv preprint arXiv:2401.16960.

[20] Yaming Yang, Zhuofeng Luo, Zhe Wang, Weigang Lu, Yiheng Lu, Ziyu Guan, Wei Zhao, and Yuanhai Lv. 2025. A translation-based heterogeneous graph neural network for multiple knowledge graphs alignment. In 2025 IEEE 41st International Conference on Data Engineering (ICDE), pages 2215–2226. IEEE.

[21] Yaming Yang, Zhe Wang, Ziyu Guan, Wei Zhao, Xinyan Huang, and Xiaofei He. 2025. Unsupervised entity alignment based on personalized discriminative rooted tree. arXiv preprint arXiv:2502.10044.

[22] Yaming Yang, Zhe Wang, Ziyu Guan, Wei Zhao, Weigang Lu, Xinyan Huang, Jiangtao Cui, and Xiaofei He. 2024. Aligning multiple knowledge graphs in a single pass. arXiv preprint arXiv:2408.00662.

[23] Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view knowledge graph embedding for entity alignment. pages 5429–5435.

[24] Yu Zhao, Yike Wu, Xiangrui Cai, Ying Zhang, Haiwei Zhang, and Xiaojie Yuan. 2023. From alignment to entailment: A unified textual entailment framework for entity alignment. pages 8795–8806.

[25] Ziyue Zhong, Meihui Zhang, Ju Fan, and Chenxiao Dou. 2022. Semantics driven embedding learning for effective entity alignment. In 2022 IEEE 38th International Conference on Data Engineering (ICDE), pages 2127–2140. IEEE.

[26] Xuhui Jiang, Yinghan Shen, Zhichao Shi, Chengjin Xu, Wei Li, Zixuan Li, Jian Guo, Huawei Shen, and Yuanzhuo" Wang. 2024. Unlocking the power of large language models for entity alignment. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 7566–7583.