Entity Profile Generation and Reasoning with LLMs for 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) involves identifying and linking equivalent entities across different knowledge graphs (KGs). While knowledge graphs provide structured information about real-world entities, only a small fraction of these entities are aligned. The entity alignment process is challenging due to heterogeneity in KGs, such as differences in structure, terminology, and attribute details. Traditional EA methods use multi-aspect entity embeddings to align entities. Although these methods perform well in certain scenarios, their effectiveness is often constrained by sparse or incomplete data in knowledge graphs and the limitations of embedding techniques. We propose **ProLEA** (**Pro**file Generation and Reasoning with **L**LMs for **E**ntity **A**lignment) an entity alignment method that combines large language models (LLMs) with entity embeddings. LLMs generate contextual profiles for entities based on their properties. Candidate entities identified by entity embedding techniques are then re-evaluated by the LLMs, using its background knowledge and the generated profile. A thresholding mechanism is introduced to resolve conflicts between LLMs predictions and embedding-based alignments. This method enhances alignment accuracy, robustness, and explainability, particularly for complex, heterogeneous knowledge graphs. Furthermore, ProLEA is a generalized framework. Its profile generation and LLM-enhanced entity alignment components can improve the performance of the existing entity alignment models.

1 Introduction

Entity alignment is the task of integrating heterogeneous knowledge among different knowledge graphs (KGs). Knowledge Graph (KG) is a popular way of storing facts about real-world entities. Traditional EA methods mainly rely on measuring the similarity of entity embeddings derived from knowledge representation learning. These methods, however, do not consider external knowledge, limiting their ability to align entities accurately, particularly in heterogeneous KGs. Recently, large language models have shown great potential in enhancing EA tasks. LLMs are trained on vast data and can provide valuable external knowledge about entities, which can improve alignment accuracy. Additionally, their strong reasoning abilities have been demonstrated in knowledge extraction and reasoning tasks.

Studies such as CHATEA (Jiang et al., 2024a) and LLMEA (Yang et al., 2024a) have explored integrating LLMs into EA. CHATEA uses LLMs for iterative reasoning and incorporates diverse data types (names, descriptions, temporal data), while LLMEA leverages entity structure and name embeddings for candidate selection, refining the final alignments through LLMs' reasoning. These approaches show the promising role of LLMs in improving EA methods. In this paper, we propose ProLEA framework that leverages large language models to generate a summarized entity profile. These profiles integrate all available properties of an entity, including its attributes, relationships, and contextual information, into a coherent and structured representation. This profile acts as a reference for evaluating potential alignments. Candidate entities are initially identified using entity embedding learning techniques, which generate an alignment shortlist based on similarity metrics. These candidates are then re-evaluated by the Large Language Model (LLM), utilizing its background knowledge and the generated entity profile to ensure precision.

We introduce a thresholding mechanism to resolve conflicts between the predictions of large language models and candidate alignments generated by the embedding module. If the LLM confirms a candidate, it is accepted as the final alignment. Otherwise, LLM generates a confidence score for its prediction. When the LLM is highly confident that the candidate is incorrect, the alignment is

Findings of the Association for Computational Linguistics: EMNLP 2025, pages 20073–20086
November 4-9, 2025

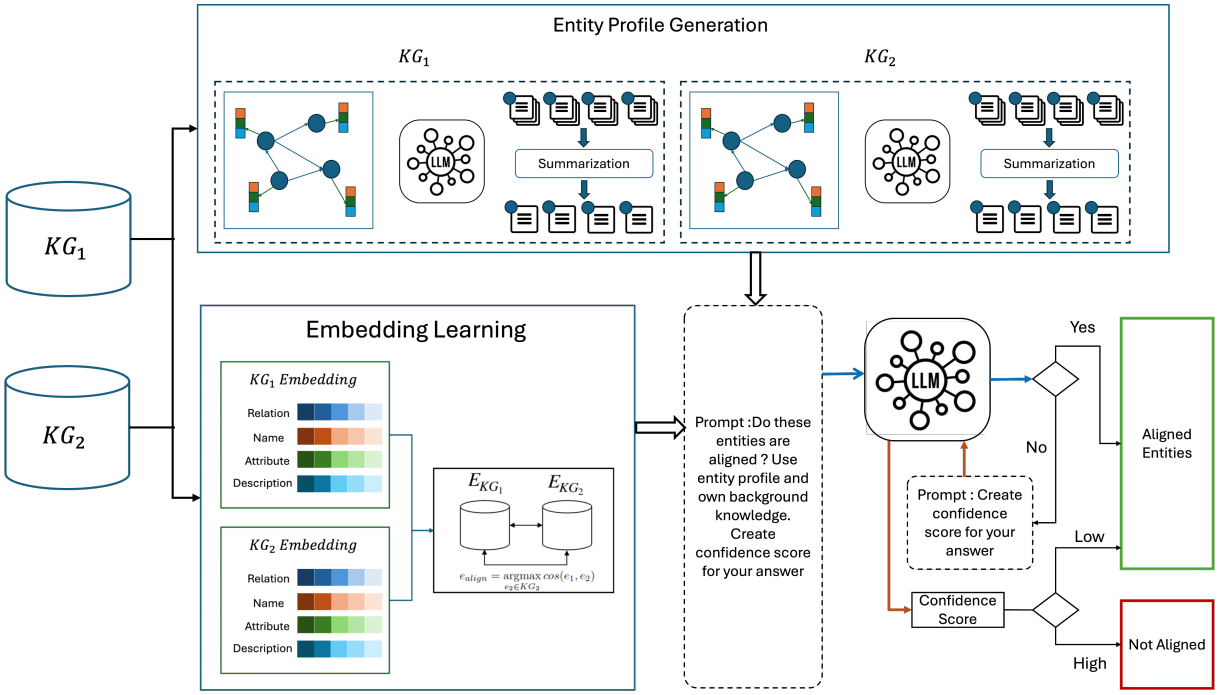

Figure 1: An overview of ProLEA.

rejected. However, if the confidence is low, indicating uncertainty, the original candidate alignment is retained. This approach ensures that only the most reliable alignments are selected, while maintaining the alignment process's accuracy, and explainability, particularly in the case of complex and heterogeneous knowledge graphs. Overall, our contributions are summarized as follows: (1) We introduce LLM-generated profiles combining attributes, relationships, and context for precise entity alignment; (2) Combine embedding-based candidate generation with LLM refinement for improved alignment accuracy; (3) Develop a thresholding mechanism for resolving conflicts between embeddings and LLM predictions based on confidence scores; (4) Improve explainability and robustness for complex knowledge graphs, balancing accuracy and adaptability through LLM-guided reasoning for reliable alignments; (5) We show that ProLEA functions as a modular post-processing component that enhances the performance of existing EA models (Jiang et al., 2024b; Tang et al., 2020) without requiring changes to their original architectures.

2 Related Works

Over the years, numerous approaches have been developed, each employing distinct methodologies to tackle the complexities of entity alignment effectively. In this section, we introduce the related work from various aspects. Translation-based methods represent entities and relations as low-dimensional vectors in a shared space, modeling relations as translations between entity embeddings (Bordes et al., 2013; Rahman and Takasu, 2018, 2020). Models like MTransE (Chen et al., 2017), BootEA (Sun et al., 2018), JAPE (Sun et al., 2017), KECG (Li et al., 2019), TransEdge (Sun et al., 2019), JarKA (Chen et al., 2020), embed entities and relations into a shared vector space to map entities across different knowledge graphs. These methods rely on translation-based knowledge graph embedding techniques to map entities across different knowledge graphs by constraining entity embeddings into a fixed distribution, facilitating cross-KG alignment. GNN based models like GCN-Align (Wang et al., 2018) and, RDGCN (Chen et al., 2022) leverage GNNs to integrate structural similarity and local graph neighborhoods, achieving comparable results. Yang et al. (Yang et al., 2025a, 2024b, 2025b) also explore GNN-based and unsupervised approaches for entity alignment. BERT-INT (Tang et al., 2020) , JSAE (Munne and Ichise, 2020) and EASAE (Munne and Ichise, 2023) employ pre-trained language models to capture semantic richness from entity descriptions, while MultiKE (Zhang et al., 2019), TEA (Zhao et al., 2023) and SDEA (Zhong et al., 2022) utilize multi-view information for alignment. Recent studies,

KG 1 (Wikidata):	KG 2 (DBpedia):
Entity Name: Albert Einstein **Description:** "Albert Einstein was a theoretical physicist known for developing the theory of relativity." **Attributes:** Date of Birth: March 14, 1879 Field of Work: Physics Notable Work: Theory of Relativity **Relationships:** Employer: Princeton University Awarded: Nobel Prize in Physics	**Entity Name:** Albert Einstein **Description:** "Albert Einstein was a renowned physicist who revolutionized our understanding of space, time, and energy." **Attributes:** Birthdate: 1879-03-14 Discipline: Theoretical Physics KnownFor: Relativity, Photoelectric Effect **Relationships:** Affiliation: Institute for Advanced Study Honors: Nobel Prize, Copley Medal
Prompt : Generate an entity profile that integrates the related information to provide a clear and concise overview of the entity	
Albert Einstein, born on March 14, 1879, was a theoretical physicist best known for developing the theory of relativity, which transformed our understanding of space and time. His groundbreaking contributions to physics include the development of the famous equation E=mc2E=mc2. Throughout his career, Einstein was affiliated with Princeton University and was awarded the Nobel Prize in Physics for his work on the photoelectric effect.	Albert Einstein, born on 1879-03-14, was a renowned physicist credited with revolutionizing the understanding of space, time, and energy. His work led to the formulation of the theory of relativity and made significant contributions to quantum mechanics. Einstein was affiliated with the Institute for Advanced Study and received numerous honors, including the Nobel Prize in Physics and the Copley Medal.
Prompt : Generate simplified and summarized version	
Albert Einstein (born 1879-03-14) was a physicist known for the theory of relativity and the equation $E=mc^2$. He won the Nobel Prize for his work on the photoelectric effect and made significant contributions to modern physics while working at Princeton University.	Albert Einstein (born 1879-03-14) was a physicist who developed the theory of relativity and contributed to quantum mechanics, fundamentally changing modern physics. He received the Nobel Prize in Physics and the Copley Medal. Einstein was affiliated with the Institute for Advanced Study.

Figure 2: Entity Profile Generation.

such as Simple-HHEA (Jiang et al., 2024b) uses structured embeddings with logical reasoning. AutoAlign (Zhang et al., 2023) leverages large language models for fully automatic KG alignment, improving semantic matching without manual supervision, while Chen et al. (Chen et al., 2024) address noisy LLM-generated annotations to enhance robust entity alignment. ChatEA (Jiang et al., 2024a) integrates LLMs for iterative reasoning to filter candidates, leveraging names, descriptions, and structures and LLMEA (Yang et al., 2024a) integrates multiple similarity features with LLM-based refinement. In contrast, our proposed ProLEA model improves alignment accuracy, robustness, and explainability by addressing modularity, data sparsity, and generalization challenges in entity alignment through LLM-driven profiling and reasoning.

3 Preliminary

In this section, we formally define the terms used in this paper and the problem as well.

Definition 1 *Knowledge Graph (KG): A knowledge graph KG = (E, R, T) , where E, R and T are the sets of entities, relations, and triples, respectively.*

Definition 2 *Relational Triples: $T \subset E \times R \times E$ is a set of relational triples representing the relations between entities, where E and R are the sets of all entities and relations respectively.*

Definition 3 *Attribute Triples: $A_T \subset E \times A \times L$ is a set of attribute triples representing the attributes of entities, where A is the set of all attributes, and each attribute $A_i \in A$ has a corresponding literal attribute value set $L_i \in L$.*

Definition 4 *Given two KGs, KG_1 and KG_2, the entity alignment problem aims to find all pairs like (e_1, e_2) where $e_1 \in KG_1$, $e_2 \in KG_2$, and e_1, e_2 represent the same real-world entity.*

Given two knowledge graphs, our objective is to jointly learn their structure, attributes, and descriptions to identify all entity pairs that refer to the same real-world entity. In our paper, we use bold lowercase letters to represent embedding vectors and bold uppercase letters to denote matrices.

4 Model Overview

The overall architecture of ProLEA is shown in Figure 1. This framework comprises three distinct components: (1) Entity Profile Generation, (2) Embedding Learning and, (3) LLM-enhanced Entity Alignment.

4.1 Entity Profile Generation (EPG)

Knowledge Graphs often face challenges such as heterogeneous representations, incomplete or inconsistent data, ambiguous descriptions, scalability issues, and cross-lingual barriers, which complicate entity alignment and usability. To address

these issues, we propose a prompt-based entity profile generation approach leveraging Large Language Models. This method synthesizes information from KGs into concise, context-rich profiles by transforming structured data such as descriptions, attributes, and relationships into unified, human-readable summaries. Acting as a semantic layer, these profiles bridge data gaps, enhance interpretability, and simplify alignment across diverse sources. For each entity, a descriptive profile is created independently for each knowledge graph. This process consists of two phases: profile generation and summarization. During the generation phase, a detailed textual description is created using all available information about the entity. In the summarization phase, this description is then condensed into a more concise version, retaining only the most relevant details.

Figure 2 provides examples of how profile generation and summarization work using two different knowledge graphs (KGs). It showcases how the process generates semantically similar profiles for the entity "Albert Einstein" from both Wikidata and DBpedia. Despite the entity being widely recognized, the two KGs contain distinct pieces of information. This highlights the potential of generating textual profiles to bridge the gap between different knowledge sources, enabling a unified representation of the entity by capturing key attributes and relationships from each graph.

4.2 Embedding Learning (EL)

The embedding learning model consists of four components, each contributing to the capture of semantic and structural information. These components are described in the following sections.

4.2.1 Name Embedding (NE)

To generate name embeddings for entities, we utilize BERT (Devlin et al., 2019) to encode entity names into dense vector representations. The BERT model transforms a name into a contextualized embedding that captures semantic and syntactic information about the entity as:

$$L_{\text{NE}} = \phi(\text{name}(e)) \tag{1}$$

where $\phi(\cdot)$ denotes the BERT-based encoder and $name(e)$ extracts the entity's name.

4.2.2 Relation Embedding (RE)

The relation embedding captures the structure of knowledge graphs (KGs). We adopt TransE (Bordes et al., 2013), which models a relation as a translation from the head entity to the tail entity. For a triple (h, r, t), the plausibility is measured by: $f_r(h, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$. The objective function L_{RE} ensures valid triples to have lower scores than corrupted ones[1]:

$$L_{RE} = \sum_{(h,r,t)\in T_r} \sum_{(h',r,t')\in T'_r} \max\big(0,\ \gamma + f_r(h,t) - f_r(h',t')\big) \tag{2}$$

where T_r and T'_r are the sets of positive and negative triples, respectively.

4.2.3 Attribute Embedding (AE)

We use TransE for attribute embedding, but unlike in relational embedding, we treat the predicate r as a translation from the entity h to its attribute value a. To encode the attribute value of entities, we employ a compositional function $\phi(a)$ where the attribute value a consists of a sequence of characters or tokens: $a = c_1, c_2, \ldots, c_l$ and define the relationship of each element in an attribute triple as $h + r \approx \phi(a)$. We compute $\phi(a)$ by summing the n-gram combinations of the attribute values, following the approach used in AttributeE (Trisedya et al., 2019).

We define the attribute embedding loss L_{AE} using a margin-based ranking loss, similar to relational embedding:

$$L_{AE} = \sum_{(h,r,a)\in T_a} \sum_{(h',r,a')\in T'_a} \max\big(0,\ \gamma + f_a(h,a) - f_a(h',a')\big) \tag{3}$$

Here, $f_a(h, a) = \|\mathbf{h} + \mathbf{r} - \phi(a)\|$. T_a and T'_a are the sets of positive and negative attribute triples, respectively[1].

4.2.4 Description Embedding (DE)

In RE, we primarily use fact triples as input. However, to address zero-shot scenarios where facts may be missing, we also incorporate embeddings derived from textual entity descriptions available within the KG. In the following equation, L_{DE} is

[1]The norm $\|\cdot\|$ can be instantiated as either the ℓ_1 or ℓ_2.

the loss function for description-based representations.

$$L_{DE} = f_{dd} + f_{dt} + f_{hd} \quad (4)$$

where $f_{dd} = \|\mathbf{h}_{\text{desc}} + \mathbf{r} - \mathbf{t}_{\text{desc}}\|$ denote the score function where both the head and tail entity representations ($\mathbf{h}_{\text{desc}}, \mathbf{t}_{\text{desc}}$) are derived from BERT-based textual descriptions. In contrast, $f_{dt} = \|\mathbf{h}_{\text{desc}} + \mathbf{r} - \mathbf{t}\|$ and $f_{hd} = \|\mathbf{h} + \mathbf{r} - \mathbf{t}_{\text{desc}}\|$ represent hybrid cases where only one entity (head or tail, respectively) uses a BERT-based representation, while the other relies on a structure-based embedding[1].

We combine the four embeddings (Name, Relational, Attribute, and Description) into an ensemble method to achieve better performance. Inspired by MultiKE (Zhang et al., 2019) model, we assign weights to each entity embedding to highlight the most important components. To compute the weight, we first calculate the average embedding of the four embeddings and then measure the deviation of each individual embedding from this average. If an embedding deviates significantly from the average, it is assigned a lower weight, as its impact on the overall alignment is considered less significant. Given an entity $e_1 \in KG_1$, we compute the similarity between e_1 and all entities $e_2 \in KG_2$ to find the aligned entity pair as:

$$e_{align} = \underset{e_2 \in KG_2}{\arg\max} \cos(e_1, e_2) \quad (5)$$

4.3 LLM-enhanced Entity Alignment (LEA)

This method utilizes a two-step process to align candidate entities with a target entity, leveraging large language models for reasoning and evaluation. In the first step, structured prompts guide the LLMs to assess the results generated by the multi-aspect embedding learning module using entity profiles and the LLM's inherent knowledge. The first prompt asks the LLMs to evaluate whether the candidate entity aligns with the target entity based on their profiles and the LLM's background knowledge. The LLM responds with [YES] or [NO]. If the response is [YES], the candidate entity is accepted as aligned. However, if the answer is [NO], the second prompt is used to generate a confidence score, quantifying the level of misalignment. The confidence score ranges from 0 to 1, where a high score indicates strong non-alignment, and a low score implies mild non-alignment. We determined the optimal confidence threshold (CT) based on empirical analysis to ensure maximum alignment accuracy. Entity pairs with confidence scores exceeding this CT are considered non-aligned, effectively filtering out less certain candidate pairs.

Table 1: Dataset Statistics.

Dataset		#Entities	#Relations	#Anchor
DBP15K (EN-FR)	EN	15,000	193	15,000
	FR	15,000	166	
DBP-WIKI	DBP	100,000	413	100.000
	WIKI	100,000	261	
ICEWS-WIKI	ICEWS	11,047	272	5,058
	WIKI	15,896	226	
ICEWS-YAGO	ICEWS	26,863	272	18,824
	YAGO	22,734	41	

Prompt 1: Entity Match Decision

> 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.

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

> If responded with `[NO]` in Prompt 1, provide a confidence score between `0` and `1` that indicates the degree of misalignment, where:
>
> - A score close to `1` indicates strong evidence that the entities are different.
> - A score close to `0` indicates uncertainty or partial overlap.

5 Experiment

In this section, we evaluate ProLEA to assess its effectiveness in entity alignment tasks. Our investigation is guided by four key research questions:

- RQ1: How effectively does ProLEA perform in entity alignment, and how does it address the limitations of existing methods?

Table 2: Evaluation results on entity prediction with LLM

Models	DBP15K(EN-FR)			DBP-WIKI			ICEWS-WIKI			ICEWS-YAGO		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
MTransE	0.247	0.577	0.360	0.281	0.520	0.363	0.021	0.158	0.068	0.012	0.084	0.040
BootEA	0.653	0.874	0.731	0.748	0.898	0.801	0.072	0.275	0.139	0.020	0.120	0.056
GCN-Align	0.411	0.772	0.530	0.494	0.756	0.590	0.046	0.184	0.093	0.017	0.085	0.038
RDGCN	0.873	0.950	0.901	0.974	0.994	0.980	0.064	0.202	0.096	0.029	0.097	0.042
Dual-AMN	0.954	0.994	0.970	0.983	0.996	0.991	0.083	0.281	0.145	0.031	0.144	0.068
TEA-GNN	-	-	-	-	-	-	0.063	0.253	0.126	0.025	0.135	0.064
BERT	0.937	0.985	0.956	0.941	0.980	0.963	0.546	0.687	0.596	0.749	0.845	0.784
FuAlign	0.936	0.988	0.955	0.980	0.991	0.986	0.257	0.570	0.361	0.326	0.604	0.423
BERT-INT	0.990	0.997	0.993	0.996	0.997	0.996	0.561	0.700	0.607	0.756	0.859	0.793
Simple-HHEA	0.959	0.995	0.972	0.975	0.991	0.988	0.720	0.872	0.754	0.847	0.915	0.870
LLMEA	0.957	0.977	0.901	-	-	-	-	-	-	-	-	-
ChatEA	0.990	**1.000**	0.995	0.995	**1.000**	**0.998**	0.880	0.945	0.912	0.935	0.955	0.944
ChatEA (GPT 4)	0.991	**1.000**	0.996	0.996	**1.000**	**0.998**	0.955	0.960	0.956	0.965	0.978	0.965
ProLEA	**0.993**	1.000	**0.997**	**0.997**	1.000	0.998	**0.964**	**0.975**	**0.972**	**0.969**	**0.981**	**0.977**

- RQ2: What are the individual contributions of ProLEA's components, and how do different LLMs impact its effectiveness?
- RQ3: Does ProLEA successfully balance alignment accuracy with computational efficiency?
- RQ4: How effectively can ProLEA enhance existing entity alignment methods as a modular post-processing step without requiring modifications to their original architectures?

5.1 Experiment Setup

In this section, we present the datasets, baselines, model configurations, and evaluation metrics used in our experiments.

5.1.1 Datasets and Implementation Setup

We conducted experiments on four entity alignment datasets (Jiang et al., 2024a), summarized in Table 1. The DBP15K (EN-FR) and DBP-WIKI datasets are relatively simple, featuring similar KG structures with comparable entity counts and aligned characteristics like fact density. In contrast, the ICEWS-WIKI and ICEWS-YAGO datasets are more complex, with significant heterogeneity in entity numbers and structures. Additionally, these datasets present unique challenges due to discrepancies between the number of anchors/seeds and entities. In our experimental setup, 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).

5.1.2 Baselines

We carefully examined existing literature and selected 12 state-of-the-art entity alignment methods that leverage diverse input features and employ various knowledge representation learning and LLM-based techniques. These include translation-based approaches such as MTransE (Chen et al., 2017) and BootEA (Sun et al., 2018); graph neural network (GNN)-based methods including GCN-Align (Wang et al., 2018), RDGCN (Chen et al., 2022), and Dual-AMN (Mao et al., 2021), TEA-GNN (Zhao et al., 2023); as well as other models such as BERT-INT (Tang et al., 2020), FuAlign (Wang et al., 2023) and Simple-HHEA (Jiang et al., 2024b). We also include LLM-based models LLMEA (Yang et al., 2024a), and ChatEA (Jiang et al., 2024a) as baselines for comparison.

5.2 Main Results

A comprehensive comparison across multiple datasets, shown in Table 2, demonstrates ProLEA consistently outperforms the state-of-the-art EA methods, effectively addressing RQ1. ProLEA showcases strong performance on diverse benchmarks such as DBP15K, DBP-WIKI, ICEWS-WIKI, and ICEWS-YAGO highlighting its robustness and scalability. By surpassing leading models like BERT-INT and ChatEA, it proves strong capability in tackling complex and heterogeneous entity alignment challenges. On the DBP15K(EN-FR) dataset, it achieves a Hits@1 score of 0.993, slightly surpassing the prior state-of-the-art BERT-INT (0.990) and matching the perfect Hits@10 score of 1.000. For the DBP-WIKI dataset, ProLEA achieves a Hits@1 score of 0.997, exceed-

ing BERT-INT's 0.996. Though various existing models have already achieved strong results on the DBP15K(EN-FR) and DBP-WIKI datasets, the true strength of ProLEA becomes apparent on the more challenging ICEWS-WIKI and ICEWS-YAGO datasets. ProLEA achieves Hits@1 scores of 0.931 and 0.969 on these datasets, outperforming ChatEA (0.880 and 0.935, respectively) by margins of 5.8% and 3.4%. While ChatEA's performance improves considerably when upgraded to GPT-4 (0.955 and 0.965), ProLEA remains highly competitive. In the entity alignment (EA) domain, even modest accuracy gains are impactful because knowledge graphs often contain millions of entities. Notably, ProLEA enhances the performance of existing models without requiring architectural entanglement or retraining, functioning as a modular add-on that maintains consistent results across a wide range of datasets. These qualities underline ProLEA's robustness, modularity, and practical advantage in real-world EA scenarios.

Table 3: Performance comparison on ICEWS-WIKI and ICEWS-YAGO datasets (CT means Confidence Threshold).

Settings	ICEWS-WIKI		ICEWS-YAGO	
	Hits@1	MRR	Hits@1	MRR
Full Model (All Components)	0.964	0.972	0.969	0.977
Entity Profile w/o Summarization	0.910	0.958	0.951	0.965
w/o Entity Profile	0.884	0.948	0.943	0.955
LLM Reasoning w/o CT	0.923	0.964	0.957	0.973
w/o LLM Reasoning	0.696	0.782	0.812	0.855

5.3 Ablation Studies

To address RQ2, we evaluate the performance of our model under different configurations on the ICEWS-WIKI and ICEWS-YAGO datasets as part of this ablation study.

5.3.1 Component Analysis

These studies aim to determine the individual benefits of components of ProLEA. As shown in Table 3, the full model (all components) achieved the highest performance on both ICEWS-WIKI and ICEWS-YAGO datasets, with Hits@1 scores of 0.964 and 0.969, and corresponding MRR scores of 0.972 and 0.977. Removing the Summarization step from the Entity Profile Generation led to a decrease in performance, with Hits@1 scores of

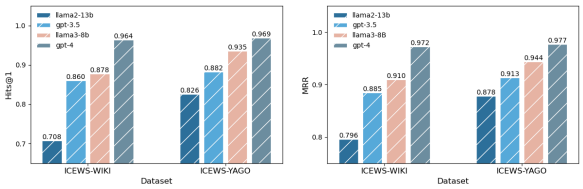

Figure 3: ProLEA's performance with different LLMs.

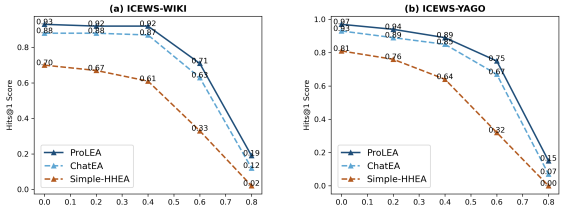

Figure 4: Performance w.r.t Embedding Noise Ratio.

0.910 and 0.951, indicating the importance of summarization for alignment and reasoning accuracy. The configuration without Entity Profile Generation led to a significant performance decline, with Hits@1 scores of 0.884 and 0.943, underscoring the critical role of this component in ensuring effective alignment and reasoning, even when the other components (embedding and LLM reasoning) remain intact. The removal of confidence threshold in LLM reasoning also caused a decrease in Hits@1 to 0.923 and 0.957, indicating that the threshold helps regulate the reasoning process, thus improving overall accuracy. However, when LLM reasoning was completely removed, the model's performance dropped significantly, with Hits@1 scores of 0.696 and 0.812, emphasizing the essential role that LLM-based reasoning plays in achieving high-quality results.

5.3.2 Performance Across Multiple LLM

We evaluate ProLEA's performance on the ICEWS-WIKI and ICEWS-YAGO datasets using a range of large language models (LLMs), as illustrated in Figure 3. This evaluation aims to assess how different LLM backbones influence the effectiveness of ProLEA's reasoning and profile-based entity alignment. Among the models tested, GPT-4 consistently achieves the highest performance, demonstrating superior reasoning capabilities and a better grasp of complex entity profiles. LLaMA 3 models rank second in performance, showing strong results but slightly trailing GPT-4, likely due to differences in training data scale and model architecture. These results highlight that ProLEA's alignment accuracy is closely tied to the capabilities of the underlying LLM. More powerful LLMs can better infer seman-

tic similarities, resolve ambiguous or sparse profile information, and make contextually informed decisions. As LLM technology continues to evolve, with improvements in contextual understanding, factual consistency, and domain specialization, ProLEA is expected to benefit proportionally, making it a future-proof and scalable solution for entity alignment across increasingly diverse and complex knowledge graphs.

5.3.3 Robustness to Embedding Noise

To evaluate ProLEA's robustness, we injected random noise into entity embeddings at varying levels, from 0% to 80%. As shown in Figure 4, Simple-HHEA relies heavily on embedding similarity, and its performance drops significantly as noise increases, indicating a strong sensitivity to the quality of input embeddings. In contrast, ChatEA shows better resilience, suggesting its capability to compensate for noisy representations. Meanwhile, ProLEA consistently maintains high accuracy across all noise levels, highlighting its strong robustness in entity alignment.

5.3.4 Confidence Threshold (CT)

We analyzed the impact of varying the confidence threshold (CT) from 0.60 to 0.95 using the random validation samples. Lower thresholds favor recall but risk false positives, while higher thresholds improve precision by filtering ambiguous matches at the cost of recall. A setting between 0.85 and 0.90 yielded the best balance across datasets. Figure 5 summarizes the Hits@1 values for ICEWS-WIKI and ICEWS-YAGO across these thresholds. We can see that performance remains high around 0.85 to 0.90, providing a clear quantitative reference for selecting a CT that balances precision and recall effectively.

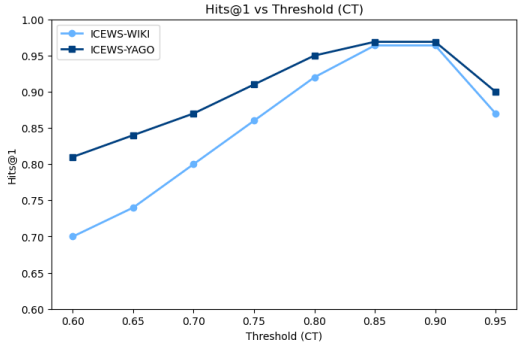

Figure 5: Hits@1 evaluation across varying CT.

5.4 Efficiency Analysis

In response to RQ3, ProLEA is designed for scenarios where accuracy in Entity Alignment (EA) is more important. According to our analysis: when using GPT-4, ProLEA processes 9,094 input tokens at an average of 122.28 ms per token, and generates 6,883 output tokens at 94.79 ms per token delivering high alignment accuracy. In comparison, with GPT-3.5, ProLEA handles 17,760 input tokens at 21.98 ms per token and produces 16,773 output tokens at 18.9 ms per token, achieving faster performance but slightly lower accuracy. These results show that ProLEA prioritizes reliability, making it especially suitable for applications like knowledge graph integration, where precision is critical. Moreover, ProLEA is compatible with different LLMs, allowing it to take advantage of future improvements in both speed and accuracy.

5.5 Extended Experiment

To assess ProLEA's flexibility as a modular post-processing step (RQ4), we integrate it with various embedding-based EA models. This experiment examines whether our model's LLM-based profiling and reasoning can augment standard embedding-based approaches without altering their original design. We adopt a two-stage evaluation strategy. In the first stage, candidate entity pairs are generated by baseline EA models such as TEA-GNN, BERT, FuAlign, BERT-INT and Simple-HHEA. In the second stage, we re-evaluate these pairs using ProLEA's profile generation and reasoning components (EPG and LEA). As shown in Figure 6, EPG and LEA consistently enhance the performance across all baselines without requiring modifications to their original architectures. Unlike embedding-only methods, whose performance can fluctuate due to graph sparsity or alignment ambiguity, using our EPG and LEA components can improve their performance across diverse datasets. Thresholding improves conflict resolution: Our proposed thresholding mechanism effectively resolves disagreements between the predictions of LLMs and embedding models. These findings confirm that ProLEA can effectively serve as a modular post-processing enhancement for existing EA methods. By re-evaluating candidate pairs through LLM-based reasoning, it consistently improves alignment accuracy, particularly in challenging scenarios involving semantic drift, noise, or missing information. ProLEA has a key advantage over methods

like ChatEA it works as a flexible add-on without changing the original model. It can improve any embedding-based method, even in noisy or sparse graphs, without adding complexity. Its confidence-based thresholding balances precision and recall, making it practical and scalable for various entity alignment tasks.

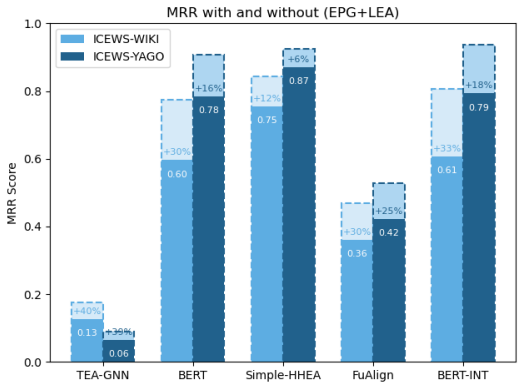

Figure 6: Evaluation results on entity prediction with and without (EPG+LEA) on MRR.

5.6 Case Study

We present two real-world examples to demonstrate ProLEA's effectiveness in handling complex entity alignment challenges such as ambiguous names and noisy or divergent attributes, where traditional embedding models often struggle. In the first case, George W. Bush (ICEWS) and George Bush (WIKI) refer to the same entity but differ in name format and attribute details. The embedding model initially ranked this pair low, likely due to name abbreviation and partial attribute overlap. ProLEA's profile generation and LLM reasoning recognized their shared role as U.S. President (2001–2009) and key historical events, correctly promoting this match to the top rank. In contrast, the embedding model wrongly ranked John Howard (Australian Prime Minister) and Howard Dean (U.S. politician) as a close match because of similar names and political context. ProLEA accurately identified their differences in country and roles, assigning a high confidence score (0.92) for non-alignment and preventing a false positive. See Appendix B for detail.

6 Conclusion

In conclusion, we propose an entity alignment method that combines large language models with entity embeddings to address challenges in heterogeneous knowledge graphs. LLMs generate contextual profiles for entities, which are re-evaluated alongside embeddings to refine candidate alignments. A thresholding mechanism resolves conflicts between LLM and embedding predictions, enhancing alignment accuracy. This approach proves effective for complex or incomplete knowledge graphs, where traditional methods often struggle. Moreover, it introduces a modular framework that enhances existing EA models through profile generation and prompt-based LLM reasoning, without requiring any architectural modifications. Extensive experiments demonstrate the model's flexibility and consistent performance gains across diverse datasets, making it practical and effective for large-scale, real-world entity alignment tasks.

Limitations

While ProLEA significantly enhances alignment accuracy through LLM-based profile reasoning, it also inherits certain limitations associated with the use of large language models. LLMs are inherently resource-intensive and have slower inference speeds compared to traditional embedding-based methods. As a result, ProLEA is best suited for applications where high alignment precision is required. However, in cases where high accuracy is not required and very low computational resources are available, the current implementation of ProLEA may face limitations. Still, for most EA tasks, high accuracy remains the most crucial factor. Moreover, although ProLEA demonstrates enhanced robustness on heterogeneous and noisy knowledge graphs, its performance can still be affected when entity information is sparse or lacks sufficient context. This limitation is not unique to ProLEA — all entity alignment models, including embedding-based and hybrid approaches, face challenges when available properties (e.g., attributes, relations, or names) are limited or ambiguous. Finally, the thresholding mechanism contributes to more reliable alignments; however, its effectiveness can vary across domains. In some cases, tuning the model for specific tasks may be necessary to balance precision and recall, although this can limit its adaptability.

Future improvements like model distillation to reduce computational costs, more efficient LLM architectures, and adaptive thresholding strategies can help address these challenges, making ProLEA more scalable and widely applicable across diverse alignment scenarios.

Ethics Statement

To the best of our knowledge, this work does not involve any discrimination, social bias, or private data. All the datasets are constructed from open-source KGs such as Wikidata, YAGO, ICEWS, and DBpedia. Therefore, we believe that our study complies with the ARR Ethics Policy.

AI Assistants

In writing this paper, we used ChatGPT (GPT-4o) to correct grammatical and keyword typos, as well as other writing issues such as inconsistent notation, while maintaining the essential content of the paper. We did not use AI tools for any of the details of our proposals and experiments, such as data generation, analysis, or evaluation methods.

References

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

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.

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

Shengyuan Chen, Qinggang Zhang, Junnan Dong, Wen Hua, Qing Li, and Xiao Huang. 2024. Entity alignment with noisy annotations from large language models. *Advances in Neural Information Processing Systems*, 37:15097–15120.

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.

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.

Xuhui Jiang, Yinghan Shen, Zhichao Shi, Chengjin Xu, Wei Li, Zixuan Li, Jian Guo, Huawei Shen, and Yuanzhuo" Wang. 2024a. 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.

Xuhui Jiang, Chengjin Xu, Yinghan Shen, Yuanzhuo Wang, Fenglong Su, Zhichao Shi, Fei Sun, Zixuan Li, Jian Guo, and Huawei Shen. 2024b. 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.

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.

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.

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.

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

Rumana Ferdous Munne, Noriki Nishida, Shanshan Liu, Narumi Tokunaga, Yuki Yamagata, Kouji Kozaki, and Yuji Matsumoto. 2025a. Zero-shot entailment learning for ontology-based biomedical annotation without explicit mentions. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 8148–8159.

Rumana Ferdous Munne, Md Mostafizur Rahman, and Yuji Matsumoto. 2025b. Lookalike audience expansion: A graph-based model with llms.

Md Mostafizur Rahman, Daisuke Kikuta, Satyen Abrol, Yu Hirate, Toyotaro Suzumura, Pablo Loyola, Takuma Ebisu, and Manoj Kondapaka. 2023. Exploring 360-degree view of customers for lookalike modeling. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3400–3404.

Md Mostafizur Rahman, Daisuke Kikuta, Yu Hirate, and Toyotaro Suzumura. 2024. Graph-based audience expansion model for marketing campaigns. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2970–2975.

Md Mostafizur Rahman and Atsuhiro Takasu. 2018. Knowledge graph embedding via entities' type mapping matrix. In *International Conference on Neural Information Processing*, pages 114–125. Springer.

Md Mostafizur Rahman and Atsuhiro Takasu. 2020. Leveraging entity-type properties in the relational context for knowledge graph embedding. *IEICE TRANSACTIONS on Information and Systems*, 103(5):958–968.

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.

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

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.

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*.

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.

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.

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.

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

Yaming Yang, Zhuofeng Luo, Zhe Wang, Weigang Lu, Yiheng Lu, Ziyu Guan, Wei Zhao, and Yuanhai Lv. 2025a. 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.

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

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

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.

Rui Zhang, Yixin Su, Bayu Distiawan Trisedya, Xiaoyan Zhao, Min Yang, Hong Cheng, and Jianzhong Qi. 2023. Autoalign: Fully automatic and effective knowledge graph alignment enabled by large language models. *IEEE Transactions on Knowledge and Data Engineering*, 36(6):2357–2371.

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.

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.

A Entity Profile Generation Prompt Templates

The entity profile generation approach employs a two-step prompting process to create informative and concise profiles from knowledge graph data. In the first phase, called generation, a prompt is designed to provide the model with all relevant entity information—such as descriptions, attributes, and relationships. The model then produces a unified and coherent profile that integrates these details into a clear, informative summary. Next, in the summarization phase, a second prompt asks the model to simplify and condense the generated profile into a brief summary of 2–3 sentences, emphasizing the most important facts while using accessible language.

Generation Prompt:

> Given the following entity information, generate a unified and coherent profile that merges the details into a single, informative summary.
>
> **Entity Name:** Albert Einstein
>
> - **Description:** Albert Einstein was a theoretical physicist known for developing the theory of relativity.
> - **Attributes:**
>   - Date of Birth: March 14, 1879
>   - Field of Work: Physics
>   - Notable Work: Theory of Relativity

Table 4: Case Study Examples from ProLEA: (Top) Correct Match via LLM Reasoning; (Bottom) False Positive Prevented via Confidence-Based Rejection.

Match Example		
Dataset	**ICEWS: George W. Bush**	**WIKI: George Bush**
Attributes	U.S. President (2001–2009), involved in War on Terror and Iraq invasion.	43rd U.S. President, term 2001–2009, Iraq War, 9/11.
Generated Profile	*"George W. Bush served as U.S. President from 2001 to 2009. He led major foreign policies including the Iraq War and post-9/11 security reforms."*	*"George Bush was the 43rd President of the U.S., known for his leadership during 9/11 and initiating the Iraq War as part of the War on Terror."*
LLM Decision	**Prompt 1 Response:** [YES] **Reasoning:** Clear semantic match in role, term, and historical context.	
Impact	Corrected misranked candidate from #5 to #1.	
Non-Match Example		
Dataset	**ICEWS: John Howard**	**WIKI: Howard Dean**
Attributes	Prime Minister of Australia (1996–2007), conservative leader.	U.S. Governor of Vermont, 2004 presidential candidate.
Generated Profile	*"John Howard was the Prime Minister of Australia known for economic reforms and foreign policy alignment with the U.S."*	*"Howard Dean is an American politician who served as Governor of Vermont and ran for U.S. President in 2004."*
LLM Decision	**Prompt 1 Response:** [NO] **Prompt 2 Result:** Confidence = 0.92 **Reasoning:** Entities differ in country, role, and political scope.	
Impact	Rejected false positive that embeddings ranked #2.	

- Employer: Princeton University
- Awarded: Nobel Prize in Physics

Summarization Prompt:

> Simplify and summarize the following entity profile into 2–3 sentences. Focus on the most important facts and use clear, accessible language.

B Case Study

Table 4. presents two illustrative case studies demonstrating ProLEA's alignment capabilities. The first example shows a correct match successfully resolved through LLM-based reasoning, despite initial ranking errors. The second example highlights the rejection of a high-ranked false positive based on confidence-informed decision-making. These cases exemplify how ProLEA leverages both semantic understanding and threshold calibration to improve entity alignment quality.

C Future Applications and Extensions

Entity alignment techniques like ProLEA can be applied in digital advertising, where audience expansion requires linking user profiles across heterogeneous platforms (Munne et al., 2025b; Rahman et al., 2023, 2024). These sources are often modeled as knowledge graphs of users, products, and interactions. They are typically noisy and sparse, which makes traditional embedding-based alignment challenging. ProLEA addresses this by generating contextual profiles with large language models and re-ranking candidate alignments using semantic reasoning, enabling more accurate user

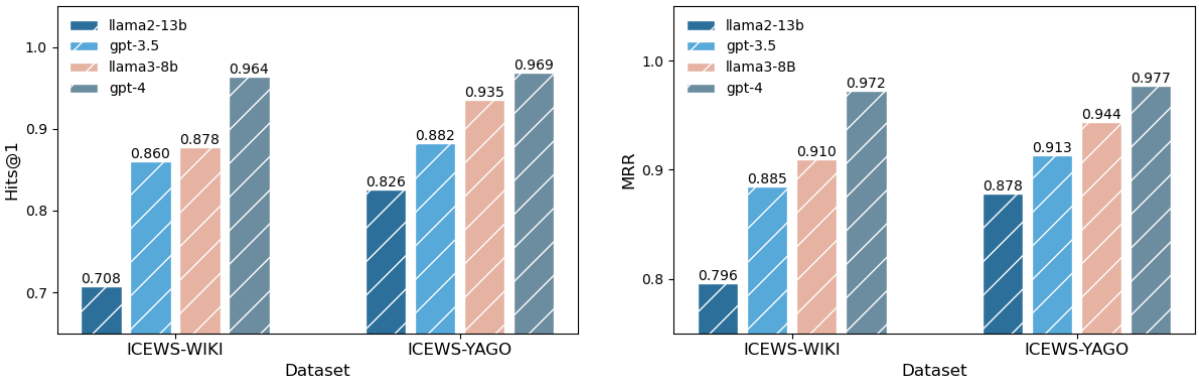

Figure 7: ProLEA's performance with different LLMs (Same as Figure 3, but shown in a larger size).

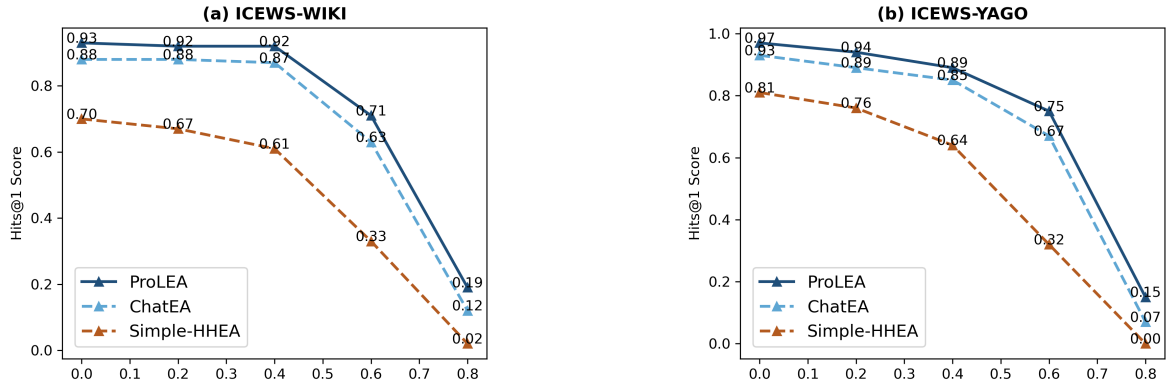

Figure 8: Performance w.r.t Embedding Noise Ratio (Same as Figure 4, but shown in a larger size).

linking and improved reach and relevance.

Similarly, in the biomedical domain faces challenges such as heterogeneous knowledge sources, reliance on costly manual curation, and the need to capture ontology concepts even when they are not explicitly mentioned in text (Munne et al., 2025a). ProLEA can align textual evidence with ontology concepts, reducing reliance on manual curation and unifying heterogeneous knowledge bases. For instance, it can map a drug's effect on a pathway to concepts like drug–target interactions, supporting scalable annotation and downstream tasks such as clinical decision support, drug discovery, and ontology-based research.

D Figures

Figures 3, 4, and 5 are provided in enlarged form in the appendix to enhance readability, as their original versions were reduced in size due to page limitations.

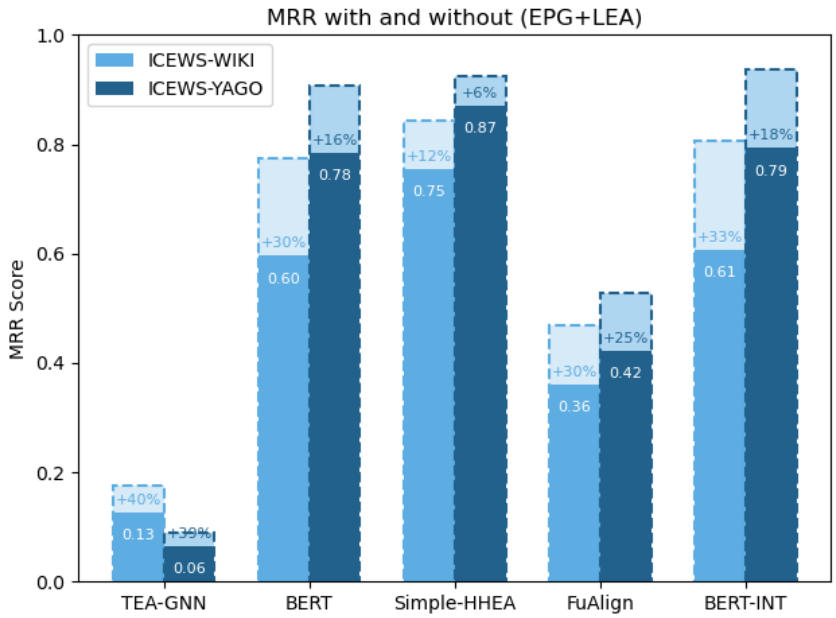

Figure 9: Evaluation results on entity prediction with and without (EPG+LEA) on MRR (Same as Figure 5, but shown in a larger size).