How do Language Models Reshape Entity Alignment? A Survey of LM-Driven EA Methods: Advances, Benchmarks, and Future

Zerui Chen[1]* **Huiming Fan**[1]* **Qianyu Wang**[1]* **Tao He**[1]
Ming Liu[1,2]† **Heng Chang**[3] **Weijiang Yu**[4] **Ze Li**[3] **Bing Qin**[1,2]
[1]Harbin Institute of Technology, Harbin, China
[2]Peng Cheng Laboratory, Shenzhen, China
[3]Huawei Inc., Shenzhen, China
[4]Sun Yat-sen University, Guangzhou, China
{zrchen,hmfan,qywang,the,mliu,qinb}@ir.hit.edu.cn
changh17@tsinghua.org.cn, weijiangyu8@gmail.com, lize23@huawei.com

Abstract

Entity alignment (EA), critical for knowledge graph (KG) integration, identifies equivalent entities across different KGs. Traditional methods often face challenges in semantic understanding and scalability. The rise of language models (LMs), particularly large language models (LLMs), has provided powerful new strategies. This paper systematically reviews LM-driven EA methods, proposing a novel taxonomy that categorizes methods in three key stages: data preparation, feature embedding, and alignment. We further summarize key benchmarks, evaluation metrics, and discuss future directions. This paper aims to provide researchers and practitioners with a clear and comprehensive understanding of how language models reshape the field of entity alignment.

1 Introduction

Large language models have demonstrated impressive performance across diverse domains, yet they remain prone to factual hallucinations—a widely recognized limitation. To address this, structured and up-to-date knowledge from knowledge graphs such as YAGO (Suchanek et al., 2007) and DBpedia (Auer et al., 2007) has emerged as a promising remedy. However, the growing demand for high-quality, domain-specific knowledge poses significant challenges, as existing KGs are often built independently with heterogeneous schemas, forming fragmented information silos. This fragmentation hinders large-scale integration and application. Consequently, as illustrated in Figure 1, entity alignment—tasked with establishing semantic correspondences among entities across distinct KGs—has regained prominence as a critical enabler for multi-source knowledge fusion and scalable KG construction, offering renewed research opportunities and practical significance.

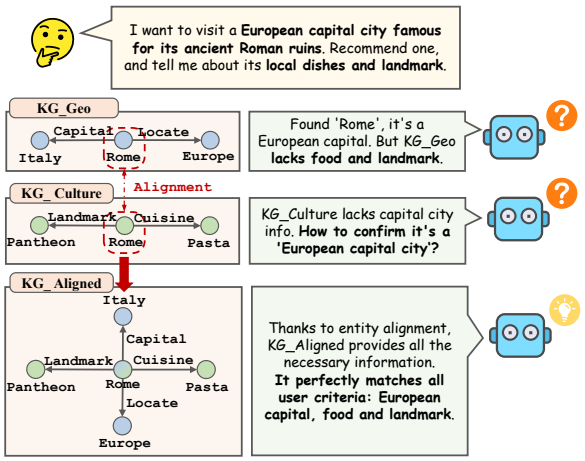

Figure 1: **Illustration of how entity alignment helps LLMs overcome information silos.** Without alignment, the LLM struggles with fragmented data from KG_Geo and KG_Culture. With alignment, KG_Aligned provides a consolidated knowledge base for a comprehensive and accurate response.

Traditional EA methods, often rooted in Knowledge Representation Learning (KRL), have explored structural information (e.g., TransE (Bordes et al., 2013a) and its extensions like IPTransE (Zhu et al., 2017), BootEA (Sun et al., 2018), and GNN-based approaches (Wu et al., 2019)) and semantic information using earlier word embeddings (e.g., Word2Vec (Church, 2017), GloVe (Pennington et al., 2014)). However, these approaches often struggle with challenges such as scalability, fully capturing deep semantic nuances, and effectively handling noisy or sparsely described entities, particularly across heterogeneous KGs. The advent of powerful language models, specifically small language models (SLMs) like BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019), and the aforementioned LLMs, has introduced transformative capabilities. Trained on massive corpora, these models encode rich linguistic, contextual, and factual knowledge, offering new avenues to overcome

* These authors contributed equally.
† Corresponding author.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 23231–23246
November 4-9, 2025 ©2025 Association for Computational Linguistics

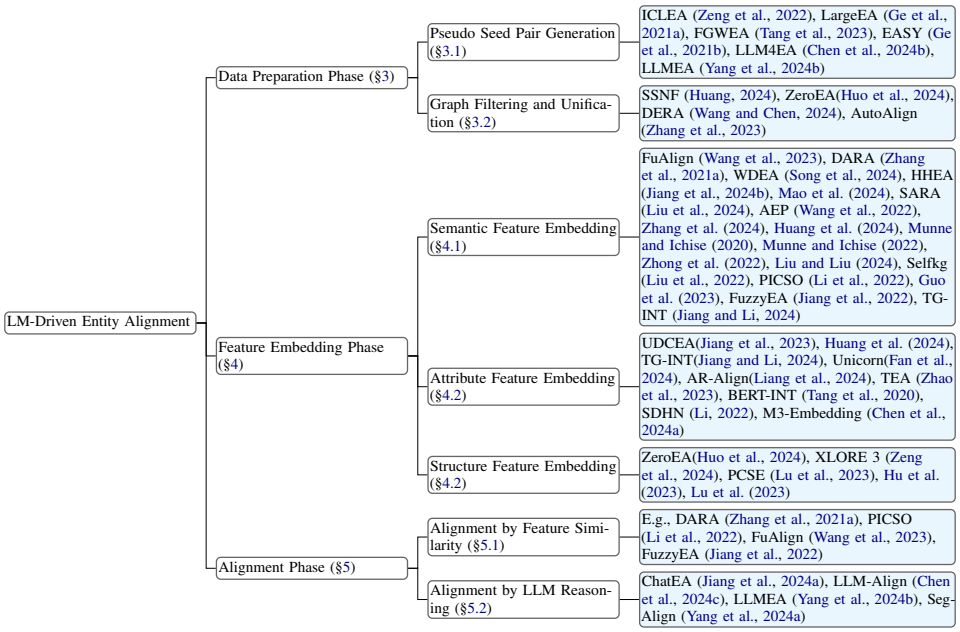

Figure 2: Classification of entity alignment models based on LMs, with a list of the most representative works.

the limitations of traditional EA techniques by better understanding entity descriptions and contexts. This survey provides a systematic and comprehensive review of EA methods driven by the advancements in language models. We introduce a novel taxonomy that classifies these contemporary approaches based on their primary role and contribution within the EA pipeline, specifically focusing on three key stages: (1) Data Preparation, which involves how LMs are used to preprocess, augment, or select data for alignment; (2) Feature Embedding, detailing how LMs generate rich entity representations crucial for similarity assessment; and (3) Alignment Strategy, covering how LM-derived features are utilized in the final matching decision process, including prompting techniques for LLMs.

While several surveys on entity alignment exist (Zhao et al., 2020; Zhang et al., 2021b; Sun et al., 2020; Fanourakis et al., 2023; Chaurasiya et al., 2022; Zhu et al., 2024), they have predominantly focused on traditional embedding models or GNN-based methods, with limited discussion on the rapidly evolving landscape of LM-driven EA. Our work specifically addresses this gap, highlighting how language models are reshaping the field. By focusing on this paradigm shift, this survey not only offers a structured understanding of current LM-based EA techniques but also underscores the synergistic potential between KGs and LMs: KGs provide structured knowledge to ground LMs, while LMs provide advanced semantic understanding to build and refine KGs through tasks like EA. We also summarize key benchmarks, evaluation metrics, and delineate promising future research directions, aiming to foster further innovation at the intersection of KGs and LMs.

Our contributions can be summarized as follows:

- **Bridge the gap in existing surveys by focusing on LM-driven EA**, reflecting the current state-of-the-art and the significant impact of these models.
- **Provide a systematic review and novel classification of EA methods driven by language models**, categorized by their application in data preparation, feature embedding, and alignment strategy stages.
- **Propose future research directions centered on forging deep synergy between KGs and LMs**, aiming to advance LM-based EA towards enhanced accuracy and reliability.

2 Preliminary

In this section, we introduce the formal definitions and notations for the entity alignment problem. We assume two knowledge graphs $\mathcal{G}$ and $\mathcal{G}'$, which are

defined as follows:

- $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$: A knowledge graph where $\mathcal{E}$ represents the set of entities, $\mathcal{R}$ denotes the set of relations, and $\mathcal{T}$ is the set of triples (e_i, r, e_j), where $e_i, e_j \in \mathcal{E}$ and $r \in \mathcal{R}$.
- $\mathcal{G}' = (\mathcal{E}', \mathcal{R}', \mathcal{T}')$: Another knowledge graph with analogous components, where $\mathcal{E}'$, $\mathcal{R}'$, and $\mathcal{T}'$ denote the set of entities, relations, and triples, respectively, in the second KG.

The objective of entity alignment is to establish a mapping $m : \mathcal{E} \to \mathcal{E}'$, such that for each entity $e \in \mathcal{E}$, there exists a corresponding aligned entity $e' \in \mathcal{E}'$. The alignment is determined based on similarity measures that capture the semantic, structural, and contextual relationships between entities from $\mathcal{G}$ and $\mathcal{G}'$.

To achieve this, an EA model is typically designed to generate vector representations for each entity. Let $f : \mathcal{E} \cup \mathcal{E}' \to \mathbb{R}^d$ denote a general embedding function that encodes each entity into a d-dimensional vector space. For each pair of entities $e \in \mathcal{E}$ and $e' \in \mathcal{E}'$, the model computes a similarity score as follows:

$$\text{score}(e, e') = s\left(f(e), f(e')\right) \tag{1}$$

where $s(\cdot, \cdot)$ is a similarity function, such as cosine similarity or negative Euclidean distance. The goal of EA is to maximize the alignment quality by ensuring that corresponding entity pairs from $\mathcal{E}$ and $\mathcal{E}'$ have higher similarity scores than non-aligned pairs. This objective can be formalized as an optimization problem:

$$\min \sum_{(e,e') \in \mathcal{M}} d\left(f(e), f(e')\right) \tag{2}$$

where $\mathcal{M}$ is the set of aligned entity pairs, and $d(\cdot, \cdot)$ is a distance metric, such as Euclidean distance or negative cosine similarity. The objective aims to minimize the distance between the embeddings of aligned entity pairs, while maximizing the distance between non-aligned pairs.

3 Data Preparation Phase

During the data preparation phase, various studies are dedicated to addressing distinct problems by leveraging language models. On the one hand, certain methodologies leverage LMs to evaluate entity pairs, consequently generating a substantial volume of pseudo-aligned entities that can facilitate subsequent model training and entity alignment tasks. On the other hand, other studies leverage certain language models for graph filtering and structural unification, aiming to facilitate feature extraction in subsequent steps.

3.1 Pseudo Seed Pair Generation

In entity alignment, pre-aligned seed pairs are typically required for model training, but these seeds are often limited and may not cover all entity types. To address this, many methods generate pseudo seed pairs to expand training data, enhance generalization, identify potential alignments, and optimize loss functions, thereby improving performance and reducing reliance on manual annotation.

Some methods leverage the encoding capabilities of language models to obtain entity features, thereby identifying pseudo seed pairs suitable for alignment. For example, ICLEA (Zeng et al., 2022) encodes entity names and descriptions separately to create pseudo-aligned pairs, while LargeEA (Ge et al., 2021a) uses BERT for entity feature encoding. Similarly, FGWEA (Tang et al., 2023) employs LaBSE (Feng et al., 2022) or SimCSE (Gao et al., 2022) to encode entity names and attributes, using the Sinkhorn algorithm (Cuturi, 2013) to establish high-confidence anchor points. Additionally, EASY (Ge et al., 2021b) combines global entity features with LMs to generate an alignment matrix and automatically create seed pairs.

The generative capabilities and instruction-following abilities of language models, especially prominent in LLMs, can also be effectively utilized to enhance the data preparation process for entity alignment. For instance, LLM4EA (Chen et al., 2024b) employs an active learning strategy to identify the most valuable entities and utilizes LMs to generate pseudo-labels for these entity pairs, thereby improving the quality of training data. Similarly, LLMEA (Yang et al., 2024b) designs tailored prompts to guide LLMs in selecting entities for alignment, generating symmetric entity pairs that are subsequently used in alignment tasks.

While these methods reduce or eliminate manual labeling and improve alignment performance, their reliance on language models for labeling may introduce noisy data, potentially affecting alignment effectiveness.

3.2 Graph Filtering and Unification

Language models, particularly LLMs, possess rich knowledge bases and exhibit exceptionally prominent zero-shot learning capabilities. Consequently,

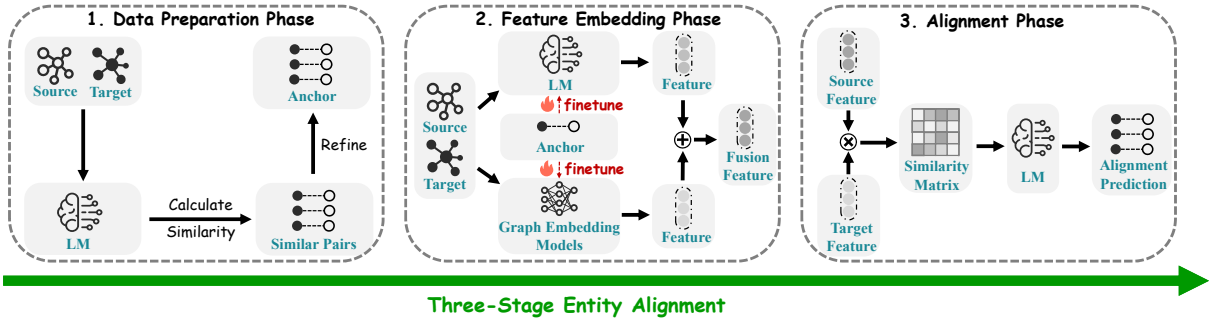

Figure 3: **Three-Stage Entity Alignment. 1. Data Preparation Phase.** LMs are used to calculate similarities between source and target entities, refining anchor pairs for alignment. **2. Feature Embedding Phase.** Fine-tuned LMs and graph embedding models generate features for source and target entities, which are fused to enhance alignment accuracy. **3. Alignment Phase.** Similarity between source and target entity representations (e.g., cosine similarity) is computed, and LMs predict alignment pairs based on the resulting similarity matrix.

some methods leverage them for graph filtering and structural unification.

For example, SSNF (Huang, 2024) exploits the zero-shot learning capability of LLMs to evaluate the importance of neighboring nodes around central entities in knowledge graphs using motifs (Yu et al., 2020), facilitating the construction of meaningful subgraphs based on node importance. ZeroEA (Huo et al., 2024) employs a Graph2Prompt module, including an LLM to transform the topological structure of the KG into textual context suitable for LM input, enabling the generation of entity features. Meanwhile, DERA (Wang and Chen, 2024) leverages the generative power of LLMs to transform entity triples into natural language descriptions, which are then embedded using BERT to create rich representations for alignment. Additionally, AutoAlign (Zhang et al., 2023) utilizes LLMs to align entity types across different knowledge graphs, simplifying predicate alignment and streamlining subsequent entity alignment tasks.

These methods rely heavily on the zero-shot capabilities of LLMs, enabling them to process entities with limited information based on the model's internal knowledge.

4 Feature Embedding Phase

During the feature embedding phase, LMs are particularly effective due to their ability to capture subtle semantic, attribute, and structure information. Recent studies have taken advantage of these strengths, integrating LMs with structured and attribute-based approaches to achieve a more accurate alignment. In the following, we categorize these advancements into three main perspectives: semantic features, attribute information, and structural embedding. Subsequently, we discuss representative methods in each category.

4.1 Semantic Feature Embedding

During the feature embedding phase, LMs are often employed to encode the semantic features of entities. These methods typically integrate contextual information, embed information using GCNs, or acquire structural information through other GNNs. Following this categorization, the subsequent sections are organized into three subsections, each detailing one of these strategies and their respective methodologies, highlighting how they complement LMs for more robust entity alignment.

4.1.1 Enriching Semantics with Neighborhood Context

To achieve robust alignment, it is essential to distinguish between two types of information LMs can process: Semantic Features, which refer to the intrinsic meaning derived from an entity's own name and description, and Contextual Cues, which provide information from its graph neighborhood, including connected entities and relations. For example, Liu and Liu (2024) proposes using the K-means algorithm to classify entities and then assess the similarity of the subgraphs in which the entities reside, while Selfkg (Liu et al., 2022) utilizes a large number of negative samples, laying the foundation for self-supervised learning-based entity alignment. To enhance encoding performance, PICSO (Li et al., 2022) integrates an entity-aware adapter with a masked self-attention mechanism, enabling the model to focus on understanding entity synonyms within specific contexts.

This design improves the model's ability to capture relationships between synonyms while preserv-

ing contextual integrity, thereby enhancing alignment accuracy.

4.1.2 Fusing LM Textual Embeddings with Graph-based Structural Embeddings

This hybrid approach combines the textual understanding of LMs with the topological awareness of models like GNNs. Typically, semantic features extracted from LMs are fused with structural features derived from GCNs to form comprehensive entity representations. For instance, FuAlign (Wang et al., 2023) employs FastText (Joulin et al., 2016) to embed entity information, while DARA (Zhang et al., 2021a) utilizes LASER (Artetxe and Schwenk, 2019) for cross-lingual entity representation. These methods focus on capturing semantic nuances through advanced embedding techniques. Furthermore, other approaches adopt BERT-like models to encode the semantic features of entities, while simultaneously utilizing GCNs to encode structural information. This dual encoding strategy is exemplified in WDEA (Song et al., 2024), HHEA (Jiang et al., 2024b), Mao et al. (2024), SARA (Liu et al., 2024), and AEP (Wang et al., 2022), which effectively combine semantic and structural representations to improve alignment accuracy.

Beyond GCNs, numerous other approaches are also employed for extracting structural information from KGs. For example, Guo et al. (2023) employs the TransH (Wang et al., 2014) model to represent relationships as hyperplanes, capturing complex relational patterns. Similarly, FuzzyEA (Jiang et al., 2022) utilizes the TransE (Bordes et al., 2013b) framework, enhanced by Dempster-Shafer evidence theory, to improve entity alignment robustness. Additionally, TG-INT (Jiang and Li, 2024) introduces QuatAE, which extracts structural features and employs a multi-view interaction mechanism to enhance alignment accuracy by considering diverse structural perspectives.

To integrate these diverse features, various techniques have been employed. For example, vector concatenation (Zhang et al., 2024) is used to merge semantic and structural embeddings into a unified representation. Weighted fusion methods (Huang et al., 2024; Munne and Ichise, 2020, 2022) dynamically adjust the contribution of each type of characteristic based on their relevance to the alignment task. Additionally, attention mechanisms (Zhong et al., 2022) are applied to selectively focus on the most informative aspects of the combined features, thereby enhancing the alignment process.

In summary, the discussed methods leverage the semantic encoding capabilities of language models and effectively integrate other features from knowledge graphs, making them a powerful and generalizable approach to entity alignment. While this fusion of embedding types is powerful, another critical source of textual information comes from an entity's structured attributes, which we discuss next.

4.2 Attribute Feature Embedding

Language models are capable of not only embedding the semantic features of entities but also capturing their attribute features, thereby enhancing the feature representation of entities. Some methods directly leverage language models to generate embeddings for these attributes, while other approaches option to further train the models to optimize their attribute embedding capabilities.

4.2.1 Direct Encoding

Some approaches leverage language models to directly encode the attribute information of entities. For instance, UDCEA (Jiang et al., 2023), Huang et al. (2024), and TG-INT (Jiang and Li, 2024) employ BERT to generate embeddings for entity attributes, integrating these embeddings with other entity information to construct features for entity alignment. Unicorn (Fan et al., 2024) employs DeBERTa (He et al., 2021) to transform serialized tuple pairs along with their attribute information into embeddings, thereby constructing features for entity alignment. AR-Align (Liang et al., 2024) utilizes LaBSE to initialize the features of entity attributes. Subsequently, these features will be fused with other features.

4.2.2 Enhanced Encoding by Training

Language models can achieve significant capability enhancements when fine-tuned on specific tasks with limited data. Consequently, certain methods leverage this training approach to improve their models' attribute embedding capabilities.

TEA (Zhao et al., 2023) reformats entity relationships and attribute triples into unified textual sequences and applies cloze-style templates as input to LMs to enhance alignment results. BERT-INT (Tang et al., 2020) enhances entity alignment by modeling intricate interactions between neighbors and attributes, integrating information from entity edges and attribute neighbors. It fine-tunes the BERT model using positive/negative sample

pairs and a pairwise margin loss.

Furthermore, numerous methods have leveraged model distillation techniques to strengthen language models' ability to embed attributes features. For instance, SDHN (Li, 2022) employs a teacher network to generate pseudo-labels that guide a student network in producing high-quality entity embeddings, thereby improving alignment performance. Similarly, M3-Embedding (Chen et al., 2024a) utilizes self-knowledge distillation to perform multi-stage training on XLM-RoBERTa, achieving multilingual, multi-function and multi-granularity entity embeddings and alignment.

These methods leverage language models to encode additional textual information of entities, particularly attributes, into entity representations, offering a simple yet effective approach. While encoding attributes provides granular detail, some of the most innovative methods take a more holistic approach to structure by translating it directly into a format LMs can process, as explored in the following section.

4.3 Verbalizing Graph Structure for LM Processing

Distinct from the embedding fusion techniques discussed in Section 4.1.2, this paradigm transforms the graph structure itself into natural language prompts or descriptions. Semantic and attribute information is relatively intuitive, while complex and implicit structures, are more challenging to capture. Nevertheless, LMs have demonstrated strong performance in extracting such complex information.

Some methods attempt to leverage rules to translate topological structures into natural language information. ZeroEA (Huo et al., 2024) employs a Graph2Prompt module that transforms the topological structure of KGs into textual context compatible with the input of the language model, following predefined rules. XLORE 3 (Zeng et al., 2024) establishes a structured concept classification system and attribute templates, integrating knowledge embedding, rule-based techniques, IR methods, and LM-based predictions.

On the other hand, diverging from fixed rule-based approaches, some methods design specialized modules for processing structural information, leveraging models for its intelligent handling. PCSE (Lu et al., 2023) utilizes contextual message passing to embed contextual information, applying an attention mechanism during feature fusion to prioritize domain-relevant information for entities. Hu et al. (2023) utilizes BERT to analyze and integrate semantic and hierarchical structures. Furthermore, Lu et al. (2023) combines attention mechanisms with the TransH model to process relational information embedded within entities.

Distinct from traditional approaches, these methods explore the utilization of language models to encode topological structures. Undoubtedly, this strategy proves effective, as structured templates and classifications enable language models to uncover a wealth of latent structural information.

5 Alignment Phase

5.1 Alignment by Feature Similarity

In the alignment phase, often building upon feature extraction methods discussed in Chapters 3 and 4, the majority of approaches initially construct pseudo-seed pairs for model training. These trained models are then utilized to extract diverse entity features. Subsequently, the similarity between the features of entities is leveraged to ascertain their alignment. This paradigm is exemplified by numerous methods. For instance, DARA fuses LASER-based semantic embeddings with GCN-based structural embeddings, with final alignment determined by finding nearest neighbors in the fused space. Similarly, BERT-INT fine-tunes BERT to produce a single, context-aware embedding, on which a straightforward similarity search is performed.

5.2 Alignment by LLM Reasoning

Large language models offer a distinct approach, often leveraging their reasoning capabilities through prompt-based interaction to evaluate the alignment potential of entities. For instance, ChatEA (Jiang et al., 2024a) translates knowledge graph structures into formats comprehensible to LLMs, compares preprocessed embeddings to identify candidate entities, and assesses alignment possibilities in a conversational manner. Similarly, LLM-Align (Chen et al., 2024c) constructs candidate entities using a language model, employs prompts based on attributes and relationships, and refines alignment results through multi-round voting. LLMEA (Yang et al., 2024b) also adopts a prompt-based approach, generating and iteratively selecting candidate entities with meticulously designed prompts. Furthermore, Seg-Align (Yang et al., 2024a) enhances cross-lingual entity alignment by identifying complex samples, utilizing zero-shot prompts,

and leveraging LLMs to achieve improved performance and efficiency. These approaches collectively demonstrate the versatility and effectiveness of LLMs in entity alignment tasks. The effectiveness of this prompt-based approach heavily depends on the prompting strategy, with key techniques including Zero-shot, Few-shot, and Chain-of-Thought. In Zero-shot Prompting, the LLM is given only instructions and entity information without any examples, testing its inherent knowledge. Few-shot Prompting enhances this by including a few examples of correct alignments to guide the model, a technique used by methods like LLMEA (Yang et al., 2024b). Finally, Chain-of-Thought Prompting elicits more complex reasoning by explicitly asking the LLM to think step-by-step (Wei et al., 2022), which improves both accuracy and interpretability.

6 Benchmarks and Metrics

6.1 Benchmarks

To evaluate EA models, existing datasets are designed to address critical challenges such as cross-lingual alignment, heterogeneous data matching, and diverse entity representations. In this survey, we utilize two widely recognized benchmark datasets, DBP15K (Sun et al., 2017) and SRPRS (Guo et al., 2019), to facilitate a comprehensive comparison of alignment methods. These datasets are chosen for their extensive language coverage and structural diversity, providing a robust basis for performance evaluation. For a broader review of other datasets used in alignment research, please refer to appendix A.

DBP15K: The DBP15K dataset consists of three cross-lingual subdatasets derived from DBpedia, containing entities aligned between English and other languages, including Chinese, Japanese, and French. Detailed statistics of the DBP15K dataset are provided in Table 5.

SRPRS: The SRPRS dataset is designed to control the degree distribution of entities, where the degree of an entity is defined as the number of relational triples associated with it. This feature allows for flexibility in experiments requiring datasets with specific structural characteristics. The statistics for SRPRS are also summarized in Table 5.

6.2 Metrics

To evaluate the performance of EA, two commonly used metrics are **Hits@k** and **Mean Reciprocal Rank**. These metrics are defined as follows:

Hits@k: Hits@k quantifies the proportion of correctly aligned entities that rank within the top k candidates, where k is typically set to 1 or 10. Each time a correctly aligned entity is found within the top k positions, the count for Hits@k increases. A higher Hits@k value reflects better model performance. These metrics are formally defined as:

$$\text{Hits@k} = \frac{\text{count}\left(\{e \in S \mid \text{rank}_e \leq k\}\right)}{\text{count}(S)} \tag{3}$$

where $\text{count}(S)$ denotes the total number of entities in the set S, rank_e is the rank of entity e, and S represents the set of all candidate aligned entities.

Mean Reciprocal Rank (MRR): Mean Reciprocal Rank measures the average reciprocal rank of all correctly aligned entities, offering insight into how early correct matches appear in the ranking list. A higher MRR value signifies better model performance. The MRR is calculated as:

$$\text{MRR} = \frac{1}{\text{count}(S)} \sum_{e \in S} \frac{1}{\text{rank}_e} \tag{4}$$

where $\text{count}(S)$ denotes the total number of entities in set S, and rank_e is the rank of entity e.

7 Evaluation and Analysis

LM-driven EA methods, evaluated on DBP15K and SRPRS datasets (Table 6 and Table 7), demonstrate significant advancements, fundamentally reshaping the field. Below we explore key insights into their multifaceted contributions.

7.1 Overall Performance Enhancement

LMs consistently achieve high Hits@1 scores across diverse language pairs and dataset characteristics within DBP15K and SRPRS. For instance, methods like FGWEA (in Data Preparation) and DERA (also leveraging LMs for data refinement) achieve near-perfect Hits@1 scores on DBP15K (e.g., FGWEA 0.987 on ZH-EN, DERA 0.994 on JA-EN), showcasing their ability to handle complex cross-lingual challenges. Similarly, TEA (in Feature Embedding) reaches 0.985 Hits@1 on the EN-FR SRPRS subset, and even direct alignment strategies using models like LLM-Align (0.983 on ZH-EN) and Seg-Align (0.988 on EN-FR SRPRS) show highly competitive results, indicating broad applicability. From this general performance, two key insights emerge:

- **LMs as Foundational Technology:** The profound semantic understanding LMs gain from vast corpora makes them a cornerstone for state-of-the-art EA. They provide a powerful baseline for capturing entity semantics, effectively overcoming semantic ambiguities that often challenged traditional, non-LM based methods.
- **Benchmark Saturation and Future Focus:** The achievement of near-perfect scores (e.g., FGWEA's Hits@1 approaching 1.000 on EN-DE SRPRS) suggests that while LMs have mastered current benchmarks, this also points to a potential saturation. Consequently, future research must pivot towards more challenging aspects like robust alignment in noisy, sparse, or highly heterogeneous KGs, and critically, towards enhancing the explainability and computational efficiency of these powerful models, rather than solely pursuing marginal gains on existing metrics.

7.2 Stage-Specific Contributions of LMs

Beyond overall performance, LMs offer versatile capabilities across the EA pipeline, leveraging different facets of their power for distinct contributions at each specific stage:

- **Data Preparation: Intelligent Data Augmenters.** LMs use deep semantic understanding and generative abilities (e.g., ICLEA, LLEMA) to create pseudo-labels (e.g., FGWEA's 0.987 Hits@1 from LM-enhanced data). This reduces manual labeling, improves data quality, bootstraps robust alignment, and aids scaling EA to larger KGs.
- **Feature Embedding: Rich, Contextualized Representations.** LMs excel at producing nuanced entity representations crucial for similarity assessment (e.g., TEA's 0.987 Hits@1). They encode textual descriptions, names, and attributes into rich embeddings (e.g., WDEA, XLORE 3) by processing full contexts. Their pre-training captures world knowledge and linguistic subtleties, yielding more discriminative features than shallower or purely structural methods.
- **Alignment Strategy: Reasoning Engines.** For direct alignment (e.g., Seg-Align's 0.988 Hits@1), LMs act as reasoning engines, using advanced reasoning, instruction-following, and few-shot learning (e.g., ChatEA). This surpasses simple embedding similarity, enabling holistic assessment of diverse evidence and paving the way for more complex and interpretable alignment via techniques like CoT prompting.

7.3 Benchmark Saturation and Practical Considerations

A significant trend emerging from the results in Table 6 and 7 is the saturation of widely-used benchmarks like DBP15K. Multiple methods consistently achieve Hits@1 scores exceeding 0.98, suggesting these datasets may no longer sufficiently differentiate top-performing models. This motivates the call for new benchmarks, as we discuss in Section 8.3.

Furthermore, practical feasibility is a crucial aspect. Fine-tuning-based approaches (e.g., BERT-INT(Tang et al., 2020)) have high training costs but fast inference, suitable for large-scale offline tasks. In contrast, API-based LLM reasoning (e.g., ChatEA(Jiang et al., 2024a)) has zero training cost but can be slow and expensive at inference, making it better for smaller or more complex alignment scenarios where high-quality reasoning is paramount.

8 Future Directions

8.1 Beyond Text: Multimodal EA

Current EA largely focuses on textual and structural data. However, real-world entities are often described through multiple modalities, including images, audio, and video. A significant future direction is Multimodal Entity Alignment, which aims to identify equivalent entities by leveraging information from these diverse sources (Zhu et al., 2023; Chen et al., 2023; Li et al., 2023; Wang et al., 2024; Xu et al., 2024a). Research in MMEA will need to tackle challenges in effectively fusing heterogeneous multimodal representations, handling missing or noisy modal data, and potentially leveraging emerging large multimodal models to achieve more comprehensive and accurate alignment across richly described entities.

8.2 Leveraging Advanced LLM Reasoning

To further boost entity alignment accuracy, particularly for complex or ambiguous cases, future work should focus on enhancing and leveraging the sophisticated reasoning capabilities of large language models. This involves moving beyond simple embedding similarity or direct prompting. Techniques such as chain-of-thought prompting (Wei et al., 2022), which encourages step-by-step reasoning, and the development of longer, more complex reasoning chains can guide LLMs to more accurately

infer entity equivalence (Chen et al., 2025), improving the discernment of subtle entity differences.

Crucially, as these advanced reasoning methods can significantly increase computational demands, their practical application necessitates concurrent efforts to enhance efficiency. Key strategies include developing more compact EA-specific models through techniques like model compression (Frantar et al., 2022) or knowledge distillation (Hinton et al., 2015; Xu et al., 2024b). Furthermore, designing efficient inference approaches, such as optimized prompting or selectively applying intensive reasoning only to the most challenging entity pairs, will be vital. Balancing the depth of LLM reasoning with computational viability is essential for realizing robust and scalable LLM-driven EA.

8.3 More Challenging and Realistic Benchmarks

The saturation observed on current benchmarks, exemplified by near-perfect scores on datasets like SRPRS, highlights a critical need for new evaluation standards. Future work must prioritize the creation and curation of more challenging datasets that genuinely test the robustness and scalability of entity alignment models. These new benchmarks should aim to incorporate key characteristics such as increased noise and sparsity to better reflect real-world KGs, which often suffer from incomplete or inconsistent information and a lack of explicit alignment cues. Furthermore, they should feature greater scale and heterogeneity, including significantly larger KGs and those with more diverse structural, semantic, or even cross-modal characteristics. Finally, such benchmarks need to present more complex alignment scenarios, such as those involving ambiguous entities, intricate n-to-m mappings, or KGs from highly specialized and diverse domains where simple surface-level features prove inadequate. Such datasets will be crucial for driving innovation beyond incremental improvements on existing metrics and fostering the development of models with true real-world applicability.

8.4 Continual and Collaborative Alignment

Real-world KGs are dynamic entities, constantly evolving with new information. This dynamism challenges static EA models, highlighting the need for continual learning paradigms that can adapt to graph updates without catastrophic forgetting of previous knowledge. Such approaches would enable models to remain accurate over time without costly, full-scale retraining. Furthermore, as data becomes increasingly decentralized, future work should explore collaborative frameworks like federated learning. These methods could facilitate privacy-preserving alignment, allowing organizations to jointly build powerful models without exposing their sensitive underlying data, which is crucial for applications in secure domains.

8.5 Parameter-Efficient Fine-Tuning for Scalable EA

The immense computational cost of fine-tuning large LMs remains a significant barrier to their widespread adoption in EA. Parameter-Efficient Fine-Tuning techniques, such as Low-Rank Adaptation (Hu et al., 2022), offer a promising solution by dramatically reducing the number of trainable parameters. Future research should focus on investigating PEFT's potential not just to lower costs but to create highly specialized yet resource-efficient alignment models.

8.6 Incorporating Human Feedback

In complex alignment scenarios with high ambiguity, automated methods often fall short. Integrating human expertise through advanced interactive frameworks is therefore essential for pushing the boundaries of model performance. Reinforcement Learning from Human Feedback (Ouyang et al., 2022) offers a powerful paradigm that can be adapted for EA. By allowing models to learn directly from expert corrections on the most challenging entity pairs, this approach can improve their reasoning capabilities and overall robustness.

9 Conclusion

This paper has delivered a comprehensive survey of entity alignment methods powered by LMs. We introduced a systematic categorization of the EA workflow into three key phases—data preparation, feature embedding, and alignment—providing an in-depth analysis of how LMs distinctively contribute to each stage. By summarizing critical benchmarks, evaluation metrics, and future directions, this work aims to serve as a valuable reference to advance EA research, fostering continued exploration and innovation at the intersection of language models and knowledge graph integration.

Limitations

Limited Coverage of Emerging Techniques. Given the rapid evolution of pre-trained language models and large language models, this survey may not fully capture the latest methods and techniques. The fast-paced development of language models and their applications in EA presents challenges in maintaining a comprehensive and up-to-date review.

Focus on General-Purpose Models. This survey primarily emphasizes general-purpose PLMs and LLMs in the context of EA tasks. While task-specific or domain-adapted models may exhibit distinctive characteristics, a detailed analysis of such models falls outside the scope of this review.

Evaluation and Benchmark Gaps. While we highlight common evaluation metrics and benchmarks for EA, some newer datasets and evaluation protocols are not included due to the scope of the review.

Lack of Supervised and Unsupervised Classification. One limitation of this review is the lack of distinction between supervised and unsupervised approaches in entity alignment methods. In this work, we primarily categorize existing methods based on the role of PLMs and LLMs across three stages of entity alignment. Consequently, we do not provide a separate classification based on supervised or unsupervised learning paradigms. As a result, the presentation of results does not explicitly differentiate between these two categories, which could provide additional insights into the effectiveness of each approach in different contexts.

Acknowledgments

The research in this article is supported by the National Science Foundation of China (U22B2059, 62276083) and the 5G Application Innovation Joint Research Institute's Project (A003).

References

Mikel Artetxe and Holger Schwenk. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610.

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *international semantic web conference*, pages 722–735. Springer.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013a. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013b. Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems*.

Deepak Chaurasiya, Anil Surisetty, Nitish Kumar, Alok Singh, Vikrant Dey, Aakarsh Malhotra, Gaurav Dhama, and Ankur Arora. 2022. Entity alignment for knowledge graphs: Progress, challenges, and empirical studies. *arXiv preprint arXiv:2205.08777*.

Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024a. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *arXiv preprint arXiv:2402.03216*.

Qiguang Chen, Libo Qin, Jinhao Liu, Dengyun Peng, Jiannan Guan, Peng Wang, Mengkang Hu, Yuhang Zhou, Te Gao, and Wanxiang Che. 2025. Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567*.

Shengyuan Chen, Qinggang Zhang, Junnan Dong, Wen Hua, Qing Li, and Xiao Huang. 2024b. Entity alignment with noisy annotations from large language models. *arXiv preprint arXiv:2405.16806*.

Xuan Chen, Tong Lu, and Zhichun Wang. 2024c. Llm-align: Utilizing large language models for entity alignment in knowledge graphs. *Preprint*, arXiv:2412.04690.

Zhuo Chen, Lingbing Guo, Yin Fang, Yichi Zhang, Jiaoyan Chen, Jeff Z Pan, Yangning Li, Huajun Chen, and Wen Zhang. 2023. Rethinking uncertainly missing and ambiguous visual modality in multi-modal entity alignment. In *International Semantic Web Conference*, pages 121–139. Springer.

Kenneth Ward Church. 2017. Word2vec. *Natural Language Engineering*, 23(1):155–162.

Marco Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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.

Ju Fan, Jianhong Tu, Guoliang Li, Peng Wang, Xiaoyong Du, Xiaofeng Jia, Song Gao, and Nan Tang. 2024. Unicorn: A unified multi-tasking matching model. *ACM SIGMOD Record*, 53(1):44–53.

Nikolaos Fanourakis, Vasilis Efthymiou, Dimitris Kotzinos, and Vassilis Christophides. 2023. Knowledge graph embedding methods for entity alignment: experimental review. *Data Mining and Knowledge Discovery*, 37(5):2070–2137.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic bert sentence embedding. *Preprint*, arXiv:2007.01852.

Elias Frantar, Saleh Ashkboos, Torsten Hoefler, and Dan Alistarh. 2022. Gptq: Accurate post-training quantization for generative pre-trained transformers. *arXiv preprint arXiv:2210.17323*.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2022. Simcse: Simple contrastive learning of sentence embeddings. *Preprint*, arXiv:2104.08821.

Congcong Ge, Xiaoze Liu, Lu Chen, Yunjun Gao, and Baihua Zheng. 2021a. Largeea: aligning entities for large-scale knowledge graphs. *Proceedings of the VLDB Endowment*, 15(2):237–245.

Congcong Ge, Xiaoze Liu, Lu Chen, Baihua Zheng, and Yunjun Gao. 2021b. Make it easy: An effective end-to-end entity alignment framework. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Dongdong Guo, Jingbo Wang, and Anna Fu. 2023. Research on entity alignment method in civil aviation equipment domain based on translation embedding. In *2023 International Symposium on Intelligent Robotics and Systems (ISoIRS)*, pages 204–214.

Lingbing Guo, Zequn Sun, and Wei Hu. 2019. Learning to exploit long-term relational dependencies in knowledge graphs. In *International conference on machine learning*, pages 2505–2514. PMLR.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: Decoding-enhanced bert with disentangled attention. *Preprint*, arXiv:2006.03654.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

Peihong Hu, Qi Ye, Weiyan Zhang, Jingping Liu, and Tong Ruan. 2023. Integration of multiple terminology bases: a multi-view alignment method using the hierarchical structure. *Bioinformatics*, 39(11):btad689.

Junbo Huang. 2024. Ssnf: Optimizing entity alignment with a novel structural and semantic neighbor filtering. In *International Conference on Knowledge Science, Engineering and Management*, pages 180–191. Springer.

Peng Huang, Meihui Zhang, Ziyue Zhong, Chengliang Chai, and Ju Fan. 2024. Representation learning for entity alignment in knowledge graph: A design space exploration. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 3462–3475.

Nan Huo, Reynold Cheng, Ben Kao, Wentao Ning, Nur Al Hasan Haldar, Xiaodong Li, Jinyang Li, Mohammad Matin Najafi, Tian Li, and Ge Qu. 2024. Zeroea: A zero-training entity alignment framework via pre-trained language model. *Proceedings of the VLDB Endowment*, 17(7):1765–1774.

Chuanyu Jiang, Yiming Qian, Lijun Chen, Yang Gu, and Xia Xie. 2023. Unsupervised deep cross-language entity alignment. *Preprint*, arXiv:2309.10598.

Jin Jiang and Mohan Li. 2024. Entity alignment based on multi-view interaction model in vulnerability knowledge graphs. page 501–516, Berlin, Heidelberg. Springer-Verlag.

Wen Jiang, Yuanna Liu, and Xinyang Deng. 2022. Fuzzy entity alignment via knowledge embedding with awareness of uncertainty measure. *Neurocomputing*, 468:97–110.

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. *arXiv preprint arXiv:2402.15048*.

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.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *Preprint*, arXiv:1607.01759.

Jia Li. 2022. A semantically driven hybrid network for unsupervised entity alignment. *ACM Transactions on Intelligent Systems and Technology*, 14:1 – 21.

Yangning Li, Jiaoyan Chen, Yinghui Li, Yuejia Xiang, Xi Chen, and Hai-Tao Zheng. 2023. Vision, deduction and alignment: An empirical study on multi-modal knowledge graph alignment. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Yangning Li, Jiaoyan Chen, Yinghui Li, Tianyu Yu, Xi Chen, and Hai-Tao Zheng. 2022. Embracing ambiguity: Improving similarity-oriented tasks with contextual synonym knowledge. *Preprint*, arXiv:2211.10997.

Yan Liang, Weishan Cai, Minghao Yang, and Yuncheng Jiang. 2024. An unsupervised multi-view contrastive learning framework with attention-based reranking

strategy for entity alignment. *Neural Networks*, 179:106583.

Ching-Hsuan Liu, Chih-Ming Chen, Jing-Kai Lou, Ming-Feng Tsai, Jiun-Lang Huang, and Chuan-Ju Wang. 2024. Sara: Semantic-assisted reinforced active learning for entity alignment. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10.

Xiao Liu, Haoyun Hong, Xinghao Wang, Zeyi Chen, Evgeny Kharlamov, Yuxiao Dong, and Jie Tang. 2022. Selfkg: Self-supervised entity alignment in knowledge graphs. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 860–870. ACM.

Yao Liu and Ye Liu. 2024. Integration of semantic and topological structural similarity comparison for entity alignment without pre-training. *Electronics*, 13(11):2036.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Chang Lu, Hongtao Zhou, and Housheng Su. 2023. Persona and contextual semantic embeddings for entity alignment. In *2023 IEEE 18th Conference on Industrial Electronics and Applications (ICIEA)*, pages 954–959.

Zhifang Mao, Chunxiao Fan, and Yuexin Wu. 2024. Self-supervised entity alignment model based on joint embedding. *Journal of Physics: Conference Series*, 2829(1):012020.

Rumana Ferdous Munne and Ryutaro Ichise. 2020. Joint entity summary and attribute embeddings for entity alignment between knowledge graphs. In *Hybrid Artificial Intelligent Systems*, pages 107–119, Cham. Springer International Publishing.

Rumana Ferdous Munne and Ryutaro Ichise. 2022. Entity alignment via summary and attribute embeddings. *Logic Journal of the IGPL*, 31(2):314–324.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Ran Song, Xiang Huang, Hao Peng, Shengxiang Gao, Zhengtao Yu, and Philip Yu. 2024. Wdea: The structure and semantic fusion with wasserstein distance for low-resource language entity alignment. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:4511–4525.

Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706.

Jian Sun, Yu Zhou, and Chengqing Zong. 2020. Dual attention network for cross-lingual entity alignment. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3190–3201.

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.

Jianheng Tang, Kangfei Zhao, and Jia Li. 2023. A fused gromov-wasserstein framework for unsupervised knowledge graph entity alignment. *Preprint*, arXiv:2305.06574.

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. *interactions*, 100:e1.

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.

Chenxu Wang, Yue Wan, Zhenhao Huang, Panpan Meng, and Pinghui Wang. 2022. Aep: Aligning knowledge graphs via embedding propagation. *Neurocomputing*, 507:130–144.

Luyao Wang, Pengnian Qi, Xigang Bao, Chunlai Zhou, and Biao Qin. 2024. Pseudo-label calibration semi-supervised multi-modal entity alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9116–9124.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI Conference on Artificial Intelligence*.

Zhichun Wang and Xuan Chen. 2024. Dera: Dense entity retrieval for entity alignment in knowledge graphs. *arXiv preprint arXiv:2408.01154*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, and Dongyan Zhao. 2019. Jointly learning entity and relation representations for entity alignment. In

Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 240–249.

Baogui Xu, Yafei Lu, Bing Su, and Xiaoran Yan. 2024a. Position-aware active learning for multi-modal entity alignment. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8215–8219. IEEE.

Xiaohan Xu, Ming Li, Chongyang Tao, Tao Shen, Reynold Cheng, Jinyang Li, Can Xu, Dacheng Tao, and Tianyi Zhou. 2024b. A survey on knowledge distillation of large language models. *arXiv preprint arXiv:2402.13116*.

Hsiu-Wei Yang, Yanyan Zou, Peng Shi, Wei Lu, Jimmy Lin, and Xu Sun. 2019. Aligning cross-lingual entities with multi-aspect information. *arXiv preprint arXiv:1910.06575*.

Linyan Yang, Jingwei Cheng, and Fu Zhang. 2024a. Advancing cross-lingual entity alignment with large language models: Tailored sample segmentation and zero-shot prompts. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 8122–8138, Miami, Florida, USA. Association for Computational Linguistics.

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

Shuo Yu, Yufan Feng, Da Zhang, Hayat Dino Bedru, Bo Xu, and Feng Xia. 2020. Motif discovery in networks: A survey. *Computer Science Review*, 37:100267.

Liu Yujia, Xu Tao, Su Zhehan, Li Xinsong, Xue Kaipeng, and Liu Yuxin. 2024. Entity alignment through joint utilization of multiple pretrained models for attribution relationship. In *2024 IEEE 11th International Conference on Cyber Security and Cloud Computing (CSCloud)*, pages 1–6. IEEE.

Kaisheng Zeng, Zhenhao Dong, Lei Hou, Yixin Cao, Minghao Hu, Jifan Yu, Xin Lv, Juanzi Li, and Ling Feng. 2022. Interactive contrastive learning for self-supervised entity alignment. *Preprint*, arXiv:2201.06225.

Kaisheng Zeng, Hailong Jin, Xin Lv, Fangwei Zhu, Lei Hou, Yi Zhang, Fan Pang, Yu Qi, Dingxiao Liu, Juanzi Li, et al. 2024. Xlore 3: A large-scale multilingual knowledge graph from heterogeneous wiki knowledge resources. *ACM Transactions on Information Systems*.

Gong Zhang, Yang Zhou, Sixing Wu, Zeru Zhang, and Dejing Dou. 2021a. Cross-lingual entity alignment with adversarial kernel embedding and adversarial knowledge translation. *Preprint*, arXiv:2104.07837.

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*.

Xin Zhang, Yu Liu, and Zhehuan Zhao. 2024. Semantics driven multi-view knowledge graph embedding for cross-lingual entity alignment. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11811–11815.

Yuanming Zhang, Tianyu Gao, Jiawei Lu, Zhenbo Cheng, and Gang Xiao. 2021b. Adaptive entity alignment for cross-lingual knowledge graph. In *Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part II 14*, pages 474–487. Springer.

Xiang Zhao, Weixin Zeng, Jiuyang Tang, Wei Wang, and Fabian M Suchanek. 2020. An experimental study of state-of-the-art entity alignment approaches. *IEEE Transactions on Knowledge and Data Engineering*, 34(6):2610–2625.

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. *Preprint*, arXiv:2305.11501.

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.

Beibei Zhu, Ruolin Wang, Junyi Wang, Fei Shao, and Kerun Wang. 2024. A survey: knowledge graph entity alignment research based on graph embedding. *Artificial Intelligence Review*, 57(9):1–58.

Bin Zhu, Meng Wu, Yunpeng Hong, Yi Chen, Bo Xie, Fei Liu, Chenyang Bu, and Weiping Ding. 2023. Mmiea: Multi-modal interaction entity alignment model for knowledge graphs. *Information Fusion*, 100:101935.

Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2017. Iterative entity alignment via joint knowledge embeddings. In *IJCAI*, volume 17, pages 4258–4264.

A Details of Benchmarks and Evaluations

We provide static of DBP15K and SRPRS in Table 5. The result on DBP15K and SRPRS is in Table 6 and Table 7.

IDS The IDS datasets are designed to evaluate entity alignment in cross-lingual knowledge graphs, focusing on more realistic KG structures by balancing the number of high-degree entities. IDS includes two scales, 15K and 100K, and spans two language pairs: English-French (EN-FR) and English-German (EN-DE). These subsets provide a controlled environment to benchmark EA methods across different dataset sizes and languages. The performance of LargeEA on IDS is presented in Table 1.

Method	$IDS15K_{EN-FR}$			$IDS15K_{EN-DE}$			$IDS100K_{EN-FR}$			$IDS100K_{EN-DE}$		
	H@1	H@5	MRR	H@1	H@5	MRR	H@1	H@5	MRR	H@1	H@5	MRR
LargeEA-$G_{EN\rightarrow L}$	88.4	92.2	0.90	89.2	93.4	0.91	83.9	87.5	0.86	85.6	89.1	0.87
LargeEA-$G_{L\rightarrow EN}$	89.9	92.9	0.91	90.8	94.2	0.92	84.7	87.8	0.86	85.8	89.2	0.87
LargeEA-$R_{EN\rightarrow L}$	88.7	91.9	0.90	89.2	94.0	0.91	84.4	88.0	0.86	83.4	86.7	0.85
LargeEA-$R_{L\rightarrow EN}$	89.8	92.7	0.91	91.1	94.9	0.93	84.3	87.5	0.86	86.4	89.6	0.88

Table 1: Performance of LargeEA on $IDS15K_{EN-FR}$, $IDS15K_{EN-DE}$, $IDS100K_{EN-FR}$, and $IDS100K_{EN-DE}$ datasets

DBP1M The DBP1M datasets are large-scale benchmarks derived from DBpedia, featuring millions of entities and relations. They simulate real-world EA scenarios by including unmatched entities and allowing knowledge graphs of varying sizes. DBP1M covers two language pairs, English-French (EN-FR) and English-German (EN-DE), providing a challenging and realistic testbed for large-scale entity alignment methods. The performance of LargeEA on DBP1M is also shown in Table 2.

Method	DBP1M		
	H@1	H@5	MRR
LargeEA-$G_{EN\rightarrow L}$	51.8	58.3	0.55
LargeEA-$G_{L\rightarrow EN}$	50.6	56.5	0.53
LargeEA-$R_{EN\rightarrow L}$	52.8	58.7	0.56
LargeEA-$R_{L\rightarrow EN}$	51.5	57.0	0.54

Table 2: Performance of LargeEA on DBP1M dataset

D-W-15K-V2 The D-W-15K-V2 dataset consists of two English knowledge graphs extracted from DBpedia and WikiData, respectively. It contains 15,000 pre-aligned entity links, representing shared entities across the two KGs. This dataset is designed to facilitate research on entity alignment between heterogeneous KGs originating from different data sources. By leveraging two widely-used English KGs with distinct schemas and data representations, D-W-15K-V2 provides a controlled yet challenging benchmark for evaluating EA methods.

Med-BBK-9K The Med-BBK-9K dataset is an industry-focused benchmark tailored for entity alignment in the domain of Chinese medical knowledge graphs. It includes 9,162 pre-aligned entity links between two KGs: one is an authoritative, human-annotated medical KG, and the other is automatically extracted from the Chinese online encyclopedia, Baidu Baike. This dataset captures the complexities of real-world data integration in the medical domain by combining high-quality human annotations with machine-extracted data.

Table 3 summarizes the performance of two state-of-the-art methods, FGWEA and DERA, on the D-W-15K-V2 and MED-BBK-9K datasets. The results highlight the strengths of these methods in handling heterogeneous and domain-specific knowledge graphs. FGWEA achieves strong performance on both datasets, particularly on D-W-15K-V2, while DERA excels in terms of precision, recall, and F1-score on both datasets, showcasing its effectiveness in real-world EA tasks.

Dataset	D-W-15K-V2			MED-BBK-9K		
	P	R	F_1	P	R	F_1
FGWEA	95.2	90.3	92.7	93.9	73.2	82.3
DERA	98.2	98.2	98.2	84.1	84.1	84.1

Table 3: Performance of FGWEA and DERA on D-W-15K-V2 and MED-BBK-9K datasets

DBP-WD consists of two knowledge graphs, sampled from DBpedia and Wikidata, with 100,000 aligned entity pairs. The dataset provides a challenging testbed by including entities with different names across the two knowledge graphs, making entity alignment tasks closer to real-world scenarios.

DBP-YG is derived from DBpedia and YAGO3, following the same structure as DBP-WD. It also contains 100,000 aligned entity pairs, offering a complementary evaluation setting by introducing entities from a different KG source.

For evaluation, both datasets use 30% of the aligned entities as seed alignments for training, while the remaining 70% are reserved for testing. This split ensures a standardized comparison for various alignment models. The performance of several alignment models, including MultiKE-SSL, AutoAlign-A, and SARA, on these datasets is summarized in Table 4.

Method	DBP-WD			DBP-YG		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
MultiKE-SSL	91.86	96.26	0.935	82.35	93.30	0.862
AutoAlign-A	88.73	96.91	-	91.27	95.62	-
SARA	-	-	-	99.1	99.9	0.99

Table 4: Performance of models on DBP-WD and DBP-YG datasets

B Potential Risks

The use of PLMs and LLMs for entity alignment introduces several potential risks that could impact data privacy and model fairness. Below, we outline the key risks associated with this process.

B.1 Privacy and Data Leakage Risks

Entity alignment often requires large-scale data collection and integration from multiple sources. If the raw data contains personally identifiable information, the risk of privacy breaches increases.

B.2 Bias and Fairness Issues

Bias is a well-known challenge in PLMs and LLMs, as these models are often pre-trained on large but imbalanced datasets. In the context of entity alignment, this could result in the preferential treatment of certain groups, languages, or regions.

Dataset	Task	Data Source	Entity	Relation	Attribute	Relation Triples	Attribute Triples
DBP15K	ZH–EN	Chinese	66,469	2,830	81,113	153,929	379,684
		English	98,125	2,317	71,173	237,674	567,755
	JA–EN	Japanese	65,744	2,043	58,582	164,373	354,619
		English	95,680	2,096	60,606	233,319	497,230
	FR–EN	French	66,858	1,379	45,547	192,191	528,665
		English	105,889	2,209	64,622	278,590	576,543
SRPRS	DBP–WD	DBpedia	15,000	253	-	38,421	-
		Wikidata	15,000	144	-	40,159	-
	DBP–YG	DBpedia	15,000	223	-	33,571	-
		YAGO3	15,000	30	-	34,660	-
	EN–FR	DBpedia	15,000	221	-	36,508	-
	EN–DE	DBpedia	15,000	222	-	33,532	-

Table 5: Statistics for DBP15K and SRPRS Datasets

Phase	Categories	Model	ZH–EN			JA–EN			FR–EN		
			Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
Data Preparation	**PLM**	ICLEA (Zeng et al., 2022)	0.884	0.972	-	0.919	0.975	-	0.986	0.999	-
		FGWEA (Tang et al., 2023)	0.987	0.997	0.991	0.991	0.998	0.994	0.998	1.000	0.999
		EASY (Ge et al., 2021b)	0.898	0.979	0.930	0.943	0.990	0.960	0.980	0.998	0.990
	LLM	LLEMA (Yang et al., 2024b)	0.898	0.923	-	0.911	0.946	-	0.957	0.977	-
		DERA (Wang and Chen, 2024)	0.985	0.997	0.990	0.994	0.999	0.996	0.996	0.999	0.997
Feature Embedding	**PLM**	WDEA (Song et al., 2024)	0.857	0.948	-	0.897	0.952	-	0.956	0.986	-
		H_{MAN} (Yang et al., 2019)	0.871	0.987	-	0.935	0.994	-	0.973	0.998	-
		AEP (Wang et al., 2022)	0.970	-	-	0.981	-	-	0.987	-	-
		FuzzyEA (Jiang et al., 2022)	0.863	0.984	0.909	0.898	0.985	0.933	0.977	0.998	0.986
		P-TRAEM (Yujia et al., 2024)	0.899	0.984	0.932	0.930	0.985	0.951	0.964	0.993	0.975
		Selfkg (Liu et al., 2022)	0.745	0.866	-	0.816	0.913	-	0.957	0.992	-
		PCSE (Lu et al., 2023)	0.880	0.983	0.920	0.931	0.992	0.955	0.973	0.998	0.983
		XLORE 3 (Zeng et al., 2024)	0.903	0.961	-	-	-	-	-	-	-
		TEA (Zhao et al., 2023)	0.935	0.982	0.950	0.939	0.978	0.950	0.987	0.996	0.990
		BERT-INT (Tang et al., 2020)	0.968	0.990	0.977	0.964	0.991	0.975	0.995	0.998	0.995
		AR-Align (Liang et al., 2024)	0.848	0.935	0.877	0.900	0.957	0.919	0.985	0.997	0.989
		ZeroEA (Huo et al., 2024)	0.985	0.993	0.991	0.982	0.995	0.989	0.998	0.999	0.998
Alignment	**LLM**	ChatEA (Jiang et al., 2024a)	0.980	-	0.984	0.985	-	0.993	-	-	-
		LLEMA (Yang et al., 2024b)	0.898	0.923	-	0.911	0.946	-	0.957	0.977	-
		Seg-Align (Yang et al., 2024a)	0.953	-	-	0.907	-	-	0.987	-	-
		LLM-Align (Chen et al., 2024c)	0.983	-	-	0.976	-	-	0.995	-	-

Table 6: Model performance comparison on DBP15K dataset

Phase	Categories	Model	EN–FR			EN–DE		
			Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
Data Preparation	**PLM**	FGWEA (Tang et al., 2023)	0.996	0.999	-	0.997	1.000	-
		EASY (Ge et al., 2021b)	0.965	0.989	0.970	0.974	0.992	0.980
Feature Embedding	**PLM**	SARA (Liu et al., 2024)	0.910	0.961	0.930	-	-	-
		TEA (Zhao et al., 2023)	0.985	0.996	0.990	0.987	0.997	0.990
		AEP (Wang et al., 2022)	0.986	-	-	0.992	-	-
Alignment	**LLM**	Seg-Align (Yang et al., 2024a)	0.988	-	-	0.982	-	-

Table 7: Model Performance Comparison on SRPRS Dataset