Journal of Intelligent Information and Communication

Research Paper

Graph-based Entity Linking with Heterogeneous Attention for Technical Documents

Marcel Barczyk[1,*] and Edmund Kapuściński[1]

[1] Faculty of Computer Science, AGH University of Science and Technology, Krakow, 30-059, Poland
*Corresponding author: marcel.ba@agh.edu.pl

Abstract. The entity linking in technical documents has specific issues, with many specialized terms, complex domain relationships, and the knowledge base often being outdated. This study proposes a high-end graph-based framework that combines heterogeneous graph attention mechanisms with deep context modeling. Using an adaptive relational reasoning approach, we combine information from the knowledge base and text to construct a comprehensive representation for stable entity disambiguation. Many experiments using large-scale, multi-domain datasets have shown that matching performance has significantly improved; compared to previous methods, accuracy and F1-score have increased by more than 5%, often exceeding 90%. The core components supporting the good results are graph attention, domain knowledge integration, and context integration; hyperparameter experiments determined stable and scalable settings. The system is general-purpose, maintaining good accuracy and completeness even in the absence of a vocabulary or ontology. This approach provides a stable and scalable foundation for the automatic entity resolution of engineering, scientific, and industrial documents. It has various applications in intelligent document analysis within knowledge management, information retrieval, and highly specialized technical fields.

Keywords: *Natural Language Processing, Entity Linking, Graph Neural Network, Technical Text Processing, Knowledge Base Integration, Contextual Representation*

Received on 04 December 2024, Accepted on 15 March 2025, Published on 29 March 2025

Introduction

Structured knowledge is gradually becoming one of the important components of various scientific research, engineering innovation, and intelligent information management systems [1]. When entities are mentioned in unstructured documents, entity linking can be used to connect them to the corresponding entries in authoritative knowledge bases of the relevant fields. This is an important component of automatic knowledge acquisition in scientific and engineering literature [2]. With rapid development, entity linking is being used in fields such as engineering design, biotechnology, and materials science for semantic search and literature mining [3]. Many studies have already utilized linguistic, semantic, and structural cues in technical documents to align text entities with knowledge in structured domains [4]. In the early days, most research on entity disambiguation used dictionary-based methods or heuristic rules. These methods are quite effective in certain cases, but they are insufficient to handle the volume and complexity of current technical corpora [5].

The accuracy and scalability of scientific text entity linking algorithms have recently been influenced by the developments in machine learning and graph-based models [6]. Deep neural network models, especially sequence and hierarchical encoders, have changed the direction of entity linking by directly extracting rich contextual and semantic patterns from data [7]. Using the aforementioned methods to integrate external knowledge graphs can further improve the analysis of ambiguous words and the collection of fine-grained relationships between domain entities [8]. Architectures based on attention mechanisms, such as self-attention and graph attention, are used to address the issues of long-distance dependency modeling and multi-hop semantic relationships in entity linking networks, and to improve the accuracy of candidate selection [9]. There are still serious issues in the technical field. These issues are caused by high lexical variability, the frequent

occurrence of domain-specific abbreviations, and the limited coverage of structured ontologies. In real-world applications, there is an increasing need for robust disambiguation capabilities [10].

Based on the above findings, this paper proposes a new technical text entity linking framework that leverages the representational flexibility and adaptability of Graph Attention Networks (GAT). The framework uses graphs to seamlessly connect domain knowledge, text features, and contextual clues to support the dynamic propagation and differentiation of information at the entity level. By learning from various data within the framework, issues such as ambiguous language, lack of context, and incomplete knowledge bases can be addressed. Propose foundational theories, innovative methods, experimental validation, and other extended applications. To support the dynamic propagation and differentiation of information at the entity level, a heterogeneous graph is constructed, seamlessly connecting domain knowledge, text features, and contextual clues. By using attention mechanism-based learning, various data within the framework are integrated to address issues of ambiguous language, lack of context, and incomplete knowledge bases. Conduct comprehensive empirical research to demonstrate that the method outperforms existing benchmark methods. Propose the following methods: fundamental theory, method innovation, experimental verification, and extended application.

Related Work

Entity Linking in Technical Domains

Entity linking in technical domains helps with the automatic retrieval, indexing, and construction of knowledge graphs [11]. For domains with limited vocabulary, such as chemistry and mechanical engineering, rule-based manually designed systems are used to complete the work. These systems use domain-specific vocabulary and pattern matching to perform the initial work [12]. These systems are not flexible enough to adapt to changes in the terminology of the research field [13]. Statistical methods based on co-occurrence analysis and probabilistic mapping have made some progress, but handling new or rare entities remains challenging [14]. To improve the robustness of the model, feature-driven machine learning is combined with contextual clues or syntactic dependencies [15]. The scientific names of these models are often inconsistent [16]. In order to improve the disambiguation capability of highly technical texts, distributed semantic representations have recently been constructed using neural network-based models [17]. To support the aforementioned improvements, large-scale annotated corpora have also been established for entities in the fields of materials science, biomedicine, and computer science [18]. These resources have facilitated the development of many applications in the past few years and supported system evaluations [19].

Graph Neural Networks for Knowledge Representation

Graph Neural Networks (GNNs) have been used to model complex relational data. GNNs are particularly suitable for knowledge-driven applications, such as many-to-many or graph-structured entities [20]. GNNs propagate and aggregate information through graph topology to understand end-to-end node and edge representations that contain local and global structural patterns [21]. Graph Convolutional Networks are the initial version of structural learning models. Subsequent models such as GraphSAGE and Graph Attention Networks have improved inductive learning and adjustable neighborhood selection [22]. Performs well in scientific domains with large-scale, heterogeneous, and frequently updated taxonomies and ontologies [23]. Entity linking using GNNs can jointly encode text segments and explicit knowledge base relations, thereby obtaining context-aware entity representations while enhancing the system's robustness to lexical ambiguity [24]. High-quality annotated graphs are needed for training, but GNNs are still not very popular [25]. Knowledge graphs are incomplete or general in low-resource technical fields [26].

Attention Mechanisms in Natural Language Processing

In natural language processing, the attention mechanism provides a new method that allows it to continuously focus on relevant parts of graphs or sequences during encoding and inference [27]. In entity linking, the attention mechanism can be used to weight the different strengths of contextual words or graph neighbor relationships, which can even be used for fine-grained disambiguation [28]. The widespread adoption of self-attention modules and improvements in the Transformer architecture have led to significant advancements in many areas

of natural language processing technology [29]. The combination model of attention mechanisms and Graph Neural Networks (GNNs) can use structured relational cues and contextual reasoning to address heterogeneity issues in scientific and engineering languages [30]. Attention mechanism models have recently been used to improve the interpretability and accuracy of technical entity linking systems in complex text environments [31].

Methodology

Construction of the Entity Linking Graph

The specialized field of technical entity linking begins with the accurate identification of potential entities and the careful construction of a heterogeneous graph that can recognize the deep semantics of language and knowledge bases. Use technical documentation to build a robust sequence labeling model to identify mentions of potential entities by leveraging syntactic and semantic features in the terminology. A high-performance candidate generation mechanism dynamically retrieves from a well-organized domain-specific knowledge base for each identified entity mention. The mechanism simultaneously considers direct lexical similarity and semantic relevance.

The process of constructing the graph is highly dependent on the technical ontology at the time. Each mentioned entity, potential entity, and related background noun or keyword in the graph is a node. Due to the vast number of sources for this information, nodes are classified and labeled in various ways according to their types. These methods include contextual modifiers, patent entities, components, or process parameters. In this topology, edges are no longer considered undirected but are assigned weights based on inferred domain-specific dependencies. The document contains co-reference, knowledge base links, and potential semantic similarity based on co-occurrence statistics of higher-order terms.

Node attributes need to be preprocessed and normalized to ensure good propagation and aggregation in the subsequent graph learning phase. The numerical parameters of text-derived features (such as position, part-of-speech distribution, and mention frequency) are normalized using min-max normalization to ensure that each dimension is within the range [0, 1]. This helps improve numerical stability during propagation. Through task-specific representation learning, continuous vectors are used in the space to represent semantic types and source indicators.

The initialization process of the node feature matrix constructed in the entity linking graph is as follows:

$$\mathbf{H}^{(0)} = \Phi([\boldsymbol{e}_t, \boldsymbol{p}_t, \boldsymbol{c}_t, \boldsymbol{s}_t]) \quad \text{Eq.(1)}$$

Connecting semantic entity embeddings, positional encodings, context aggregation vectors, and symbolic classification features constitutes each initial node embedding. Fusion is performed through domain adaptation transformation.

Edges achieve heterogeneous relationships through composite adjacency functions:

$$A_{ij} = \sum_{r \in \mathcal{R}} \alpha_{ij}^{(r)} \cdot \chi^{(r)}(i,j) \quad \text{Eq.(2)}$$

In this formulation, for each edge (i, j), multiple relation types r contribute weighted evidence through learned attention coefficients $\alpha_{ij}^{(r)}$, modulated by an indicator function that encodes the existence of each relation.

The feature matrix and the adjacency matrix have both been orthogonally normalized to ensure that various types of edges and nodes are comparable:

$$\tilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}};\ \tilde{\mathbf{H}}^{(0)} = \frac{\mathbf{H}^{(0)} - \mu}{\sigma} \quad \text{Eq.(3)}$$

Here, $\mathbf{D}$ is the node degree diagonal matrix, μ and σ denote the mean and standard deviation vectors of the initial feature matrix, ensuring all information channels are balanced prior to subsequent graph learning.

In this architectural design, the heterogeneous entity linking graph has multiple layers, as shown in Figure 1. Mentions and candidate entities are nodes in the graph, connected by various types of relationship edges. These edges not only contain textual proximity and co-reference information but also diverse cross-modal knowledge

from domain ontologies. In order to achieve more complex graph neural network representations downstream, provide a stream that includes context-aware, structural, and semantic information.

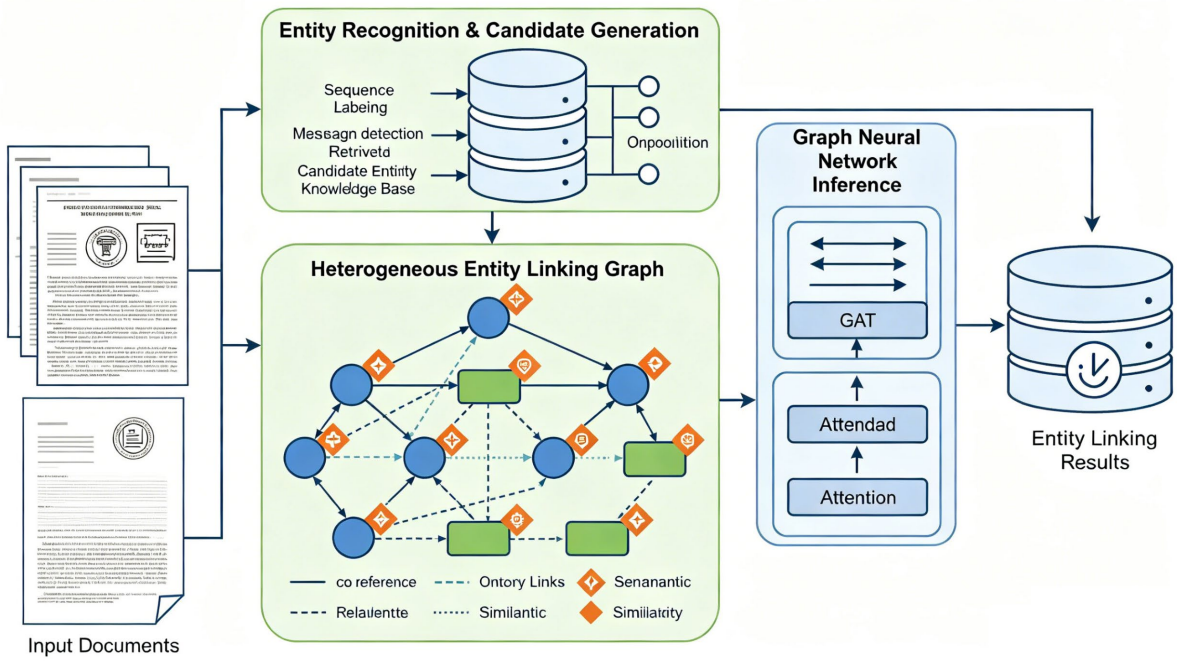

Figure 1. System Architecture of the Proposed Method.

GAT-Based Entity Representation and Linking

The method is based on a heterogeneous entity linking graph and a multi-layer Graph Attention Network (GAT). The goal is to address the complex dependency structures and contextual differences in technical corpora. Whether it is a context anchor, a mentioned entity, or a candidate entity, each node in the graph is iteratively updated based on the features of its neighbors. The attention mechanism is adjusted based on edge semantics and node types. GAT is used to address heterogeneous edge relationships and dynamically select and propagate data in the graph.

The attention mechanism uses specific combinations of structure and content to determine the weights of neighboring nodes:

$$\alpha_{ij} = \frac{\exp\left(\sigma\left(\mathbf{a}^{\top}\left[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j \| \mathbf{r}_{ij}\right]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\sigma\left(\mathbf{a}^{\top}\left[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_k \| \mathbf{r}_{ik}\right]\right)\right)} \quad \text{Eq.(4)}$$

where $\| \cdot \|$ refers to the concatenation operator, $\mathbf{W}$ is a learnable weight matrix, σ is a LeakyReLU nonlinearity, and $\mathbf{r}_{ij}$ encodes the relation type between i and j.

Update the writing on each attention head node:

$$\mathbf{h}_i' = \varphi\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}\mathbf{U}\mathbf{h}_j\right) \quad \text{Eq.(5)}$$

where φ denotes a nonlinear activation function such as ELU, and $\mathbf{U}$ is a transformation matrix acting on each neighbor's representation.

Multi-head attention enhances expressive capability by combining the outputs of different independent attention mechanisms and then propagating them. The combination update for each node in all K heads can be written as:

$$\mathbf{h}_i^{(l+1)} = \|_{k=1}^{K}\varphi\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}^{(k)}\mathbf{U}^{(k)}\mathbf{h}_j^{(l)}\right) \quad \text{Eq.(6)}$$

The aforementioned organization can acquire a large amount of structural and semantic dependencies to address the diversity and variety of technical vocabulary.

Multiple GAT layers are used for iterative message passing, enhancing node representations through progressively more abstract relational reasoning. After passing through L layers, the output nodes can be used to resolve local and global knowledge disambiguation issues.

Based on the bilinear similarity function, a candidate selection function is set up for the entity linking task, which is based on the entity mention and the final node embedding of the candidate entity. The decision for entity linking depends on the bilinear similarity function:

$$\text{score}(m, c) = \mathbf{h}_m^{(L)\top} \mathbf{M} \mathbf{h}_c^{(L)} \quad \text{Eq.(7)}$$

Here, $\mathbf{M}$ is a learnable projection matrix, and the resulting score determines the confidence of linking mention m to candidate c.

In the GAT-based reasoning process, the model continuously adjusts the text context and structures the strength of the connections between ontological and surface language cues. Help the system better learn complex backgrounds and domain knowledge to identify when subtle distinctions in terminology correspond to different physical or conceptual entities in highly specialized engineering and scientific fields.

The GAT mechanism in this paper can be explained through attention distribution, which inductively generalizes new mentions of entities. In order to improve the transparency of expert involvement in the technical workflow, post-hoc audits can be conducted on the aforementioned content to determine which contextual nodes or ontology relationships have a greater impact on making specific linking decisions.

Contextual Feature Integration

Contextual semantic signals must be integrated into the entity linking framework, especially in technical documents, where entity disambiguation often relies on complex linguistic cues and potential document structures. By embedding local and global linguistic contexts, we can construct rich contextual representations. Combine this information with knowledge based on structured graphs to establish stable links under ambiguous conditions.

The deep contextual encoder encodes each mentioned entity to extract its contextual embedding. This includes higher-level discourse structures as well as windowed word sequences from surrounding sentences, capturing the broad dependencies of the topic. Formally, the context embedding for entity mention m can be defined as

$$\mathbf{c}_m = \Psi\big(\{\mathbf{e}_w \mid w \in \mathcal{W}_m\}, \Gamma(d_m)\big) \quad \text{Eq.(8)}$$

where $\{\mathbf{e}_w\}$ are pre-trained or fine-tuned token embeddings for words in the local window $\mathcal{W}_m$, and $\Gamma(d_m)$ encodes discourse-level signals from document section d_m (such as paragraph role, subsection type, or contextual intent).

To jointly leverage knowledge-driven and context-driven information, we design a fusion layer that synergistically combines the node representations produced by GAT with the semantic context vectors. This layer employs a gated integration approach to adaptively balance structured and contextual signals according to their predictive salience for the linking task. The global feature vector for a given candidate node c after fusion is

$$\mathbf{z}_c = \delta\big(\lambda_c \cdot \mathbf{h}_c^{(L)} + (1 - \lambda_c) \cdot \mathbf{c}_c\big) \quad \text{Eq.(9)}$$

where $\mathbf{h}_c{}^{(L)}$ denotes the node's final GAT embedding, $\mathbf{c}_c$ is the contextual embedding, λ_c is a learnable or attention-derived fusion gate controlling the relative influence, and δ represents a nonlinear transformation with layer normalization to maintain numerical stability and interfeature coherence.

In addition to the aforementioned functions, the dynamic information path adjustment mode can also address issues of local ambiguity and insufficient document details. During the training process, backpropagation altered the context encoder and the graph representation module. This enables the model to find the best integration path suitable for the semantic complexity of the technical corpus.

Graph-driven reasoning and rich parameterized contextual semantics directly address the various differences in scientific language. Adding integrated modules can significantly improve the recall and precision of entity linking

in benchmark datasets. This is particularly applicable to rare words, subtle synonyms, or context-sensitive meaning changes, which can pose challenges for pattern-based or knowledge-based algorithms.

Experimental Setup

Datasets and Evaluation Metrics

These technical corpora aim to cover various organized and unorganized industrial and scientific texts, which form the empirical basis of this study. These datasets contain annotated entities from patent databases, engineering guides, and academic papers. Over 115,000 mentions and millions of pairs come from different fields (such as electronic component design, industrial automation, and new materials research). In order to conduct the evaluation, a thorough expert review and verification process must be undertaken to ensure that all mentions are accurately linked to their corresponding real entities.

Figure 2 shows the end-to-end experimental workflow and describes the processes of sourcing, filtering, annotating, and partitioning the raw data. The two components of this process are entity mention recognition and candidate set construction. Domain ontologies and lexical variants will both be part of it. Due to the scalability and adaptability of the pipeline, various technical documents can be handled in practice. The workflow diagram shows the impact of each major stage on data volume and entity diversity.

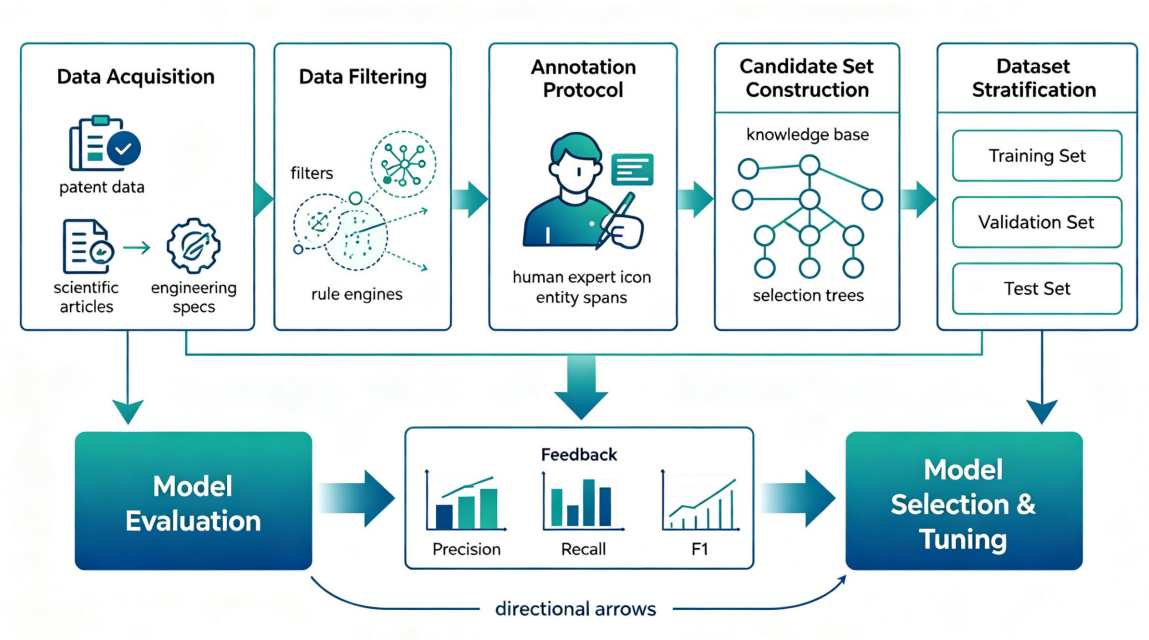

Figure 2. Experimental Workflow for Technical Entity Linking

There are three reasons for choosing these metrics to evaluate the link model: technical accuracy, etc. Precision is the percentage of correctly linked entities in each entity link prediction:

$$\text{Precision} = \frac{|\ \text{True Positives}\ |}{|\ \text{True Positives}| + |\text{False Positives}\ |} \qquad \text{Eq.(10)}$$

The ratio of correctly linked entities to all valid references in the corpus is called recall:

$$\text{Recall} = \frac{|\ \text{True Positives}\ |}{|\ \text{True Positives}| + |\text{False Negatives}\ |} \qquad \text{Eq.(11)}$$

The F1 score is the harmonic mean of precision and recall, used to address class imbalance and varying degrees of ambiguity in the data.

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad \text{Eq.(12)}$$

https://doi.org/10.64972/jiic.2025v3.219p6s:69-82

In addition to the aforementioned traditional indicators, the analytical framework also categorizes them based on time and environment to demonstrate changes in high-demand technological environments. The aforementioned indicators guided the final models and architectures chosen for all these industrial applications.

Experimental Implementation

All experiments were conducted in the same computing environment, using high-performance NVIDIA A100 Tensor Core GPUs, and the PyTorch backend optimized for sparse matrix operations and large-scale graph parallelization. By using Bayesian optimization to modify the model's hyperparameters on the validation set, we aim to stabilize convergence and reduce overfitting during cross-domain transfer.

Dynamic batching runs training instances each time to maximize hardware usage and ensure strict separation of training, validation, and test partitions. Select the test set, which will use stratified sampling to cover all types of document structures and rare terms. Improve generalization ability by using a stochastic weight averaging scheme, with early stopping conditions based on the moving average of validation loss.

The training objective is a comprehensive loss function that includes entity link classification and structural consistency penalties. Letting the predicted linkage distribution for mention m to candidate c be p_{mc} and the ground truth be y_{mc}, the total loss is given by

$$L_{\text{total}} = -\sum_{m}\sum_{c} y_{mc}\log\, p_{mc} + \beta \sum_{i,j} \omega_{ij} \left\| \mathbf{h}_i^{(\text{final})} - \mathbf{h}_j^{(\text{final})} \right\|_2^2 \qquad \text{Eq.(13)}$$

For each dataset, the weights ω_{ij} and the regularization parameter β have been adjusted. Force the embedding similarity between nodes connected by high-confidence ontology edges.

Through five independent runs using different random seeds, the results will be the average performance. To ensure reproducibility and statistical reliability, these results will include monitored standard deviations.

Baseline Methods for Comparison

In order to ensure the scientific validity of the comparison, baseline methods with diverse architectures will be selected, as these methods have performed excellently in recent technical entity linking research. Includes classic feature-based linking models, deep neural mention-entity matching networks, and advanced graph-enhanced methods, but does not use attention-based propagation. Each baseline was thoroughly re-implemented and retrained on the benchmark dataset, using the same preprocessing, candidate construction, and evaluation protocols. These protocols are the same as the proposed methods.

A scoring function can be used to create a comparison system for the comparison and grouping of all statistical data. The evaluation criteria for the summary ranking are

$$\mathcal{S}_{\text{rank}} = \frac{1}{N}\sum_{d=1}^{N} \eta_d(\gamma_1 \cdot \text{F1}_d + \gamma_2 \cdot \text{Precision}_d + \gamma_3 \cdot \text{Recall}_d) \qquad \text{Eq.(14)}$$

where N is the total number of dataset partitions, $\gamma_1, \gamma_2, \gamma_3$ are task-dependent weighting coefficients summing to one, and η_d reflects the relative technical difficulty for partition d (such as ambiguity density or ontology sparsity).

For each slice of the dataset and method, a comprehensive ablation study and error decomposition analysis will be conducted to provide detailed information on the overall system performance, as well as the specific reasons for the technical improvements brought by the new attention-based graph structure.

Results and Discussion

Quantitative Results and Analysis

The performance of the improved entity linking method has seen some improvement. After comparing with a strong baseline model, as shown in Figure 3(a), the accuracy results indicate that the model can still correctly select entity references even in the presence of a large number of overlapping and ambiguous terms. Combining

semantic structure with context-aware attention mechanisms is feasible. This helps to address references that cannot be resolved through surface lexical overlap in the corpus.

As shown in Figure 3(b), both the feature-based method and the traditional neural baseline significantly reduced the number of false positives. The aforementioned improvements are mainly attributed to the system's ability to filter candidates through dynamically weighted graph relationships. In highly specialized technical texts, the ability to reject false or contextually irrelevant matches has been enhanced.

As shown in Figure 3(c), the recall rate of authoritative entities is relatively high; the model can not only retrieve frequently occurring technical terms but also rare and frequently mentioned technical terms. In the entity linking graph, propagate context and ontology signals to achieve comprehensive coverage while reducing losses caused by overly aggressive candidate pruning.

As shown in Figure 3(d), the F1-score results of the solutions are similar. A high and stable F1 score indicates that there is a good balance between precision and recall across various dataset domains and partitions, with an increase in one dimension not being offset by a decrease in the other dimension. The following areas contain many complex documents and interrelated entities.

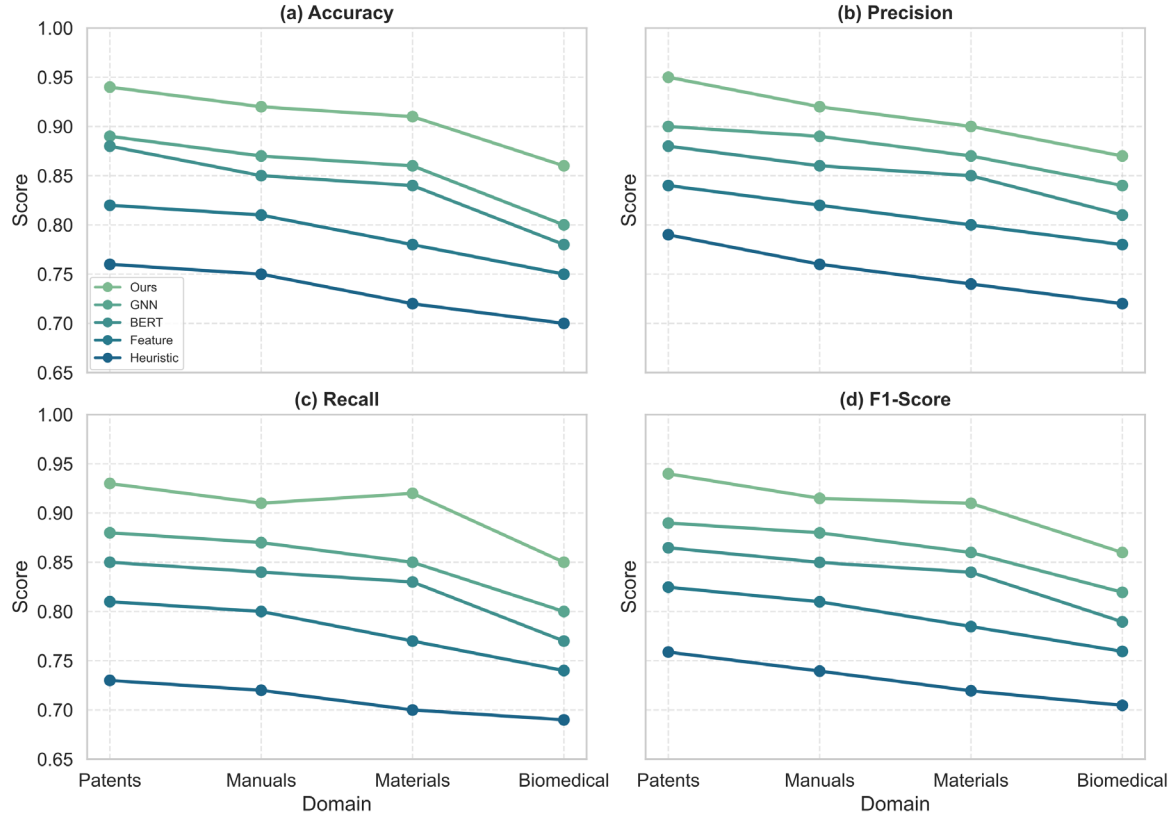

Figure 3. Comparison of Model Performance: (a) Accuracy; (b) Precision; (c) Recall; (d) F1-Score.

In performance experiments with a larger candidate pool and increased ambiguity, the model demonstrates greater scalability and stability. In the context of increasing problem complexity, the stability of the model's accuracy and recall rates indicates that it can be widely applied. Combine local text cues with the global knowledge graph to make reasonable entity linking decisions.

Based on the aforementioned metric-based insights, the combination of graph reasoning and context-sensitive encoding is beneficial for the technical entity linking task. The model not only improves the primary metrics but also establishes stable and diverse advantages at the fine-grained technical metric level. This sets a new high standard for the analysis of industrial and scientific entities.

Ablation Study

Each part of the model is dismantled, and then an ablation study is conducted to evaluate the contribution of each part to the model. To ensure the fairness and clarity of the analysis, it is necessary to sequentially remove or modify the explicit context integration in the Graph Attention Network (GAT), external knowledge bases, and experimental design, while keeping other parts unchanged.

Figure 4(a) depicts the impact of the graph reasoning layer. When a simple mean aggregation module is used instead of GAT, all performance metrics show a significant decline, with the F1-score and recall rate being particularly affected. Using an adaptive attention mechanism with heterogeneous edge semantics in technical texts is necessary. Extremely fine-grained neighborhood information weighting will be used to improve entity resolution for overlapping or nested mentions.

Figure 4(b) shows the experimental results without external knowledge base features. The lack of structured domain knowledge leads to a significant decline in the accuracy and precision of infrequent or irregular items in the system, while the basic context inference for high-frequency words remains relatively effective. Knowledge-driven context propagation helps improve the robustness of predictions under changes in terminology and domain.

The role of explicit context modeling is shown in Figure 4(c). Removing contextual embeddings from the feature set will reduce the model's recall rate, and there will be a large number of unresolved or incorrect links in text sections with ambiguous expressions or sparse co-reference chains. Context-sensitive encoding is necessary to connect local syntactic cues in technical narratives with the higher-level structure of discourse.

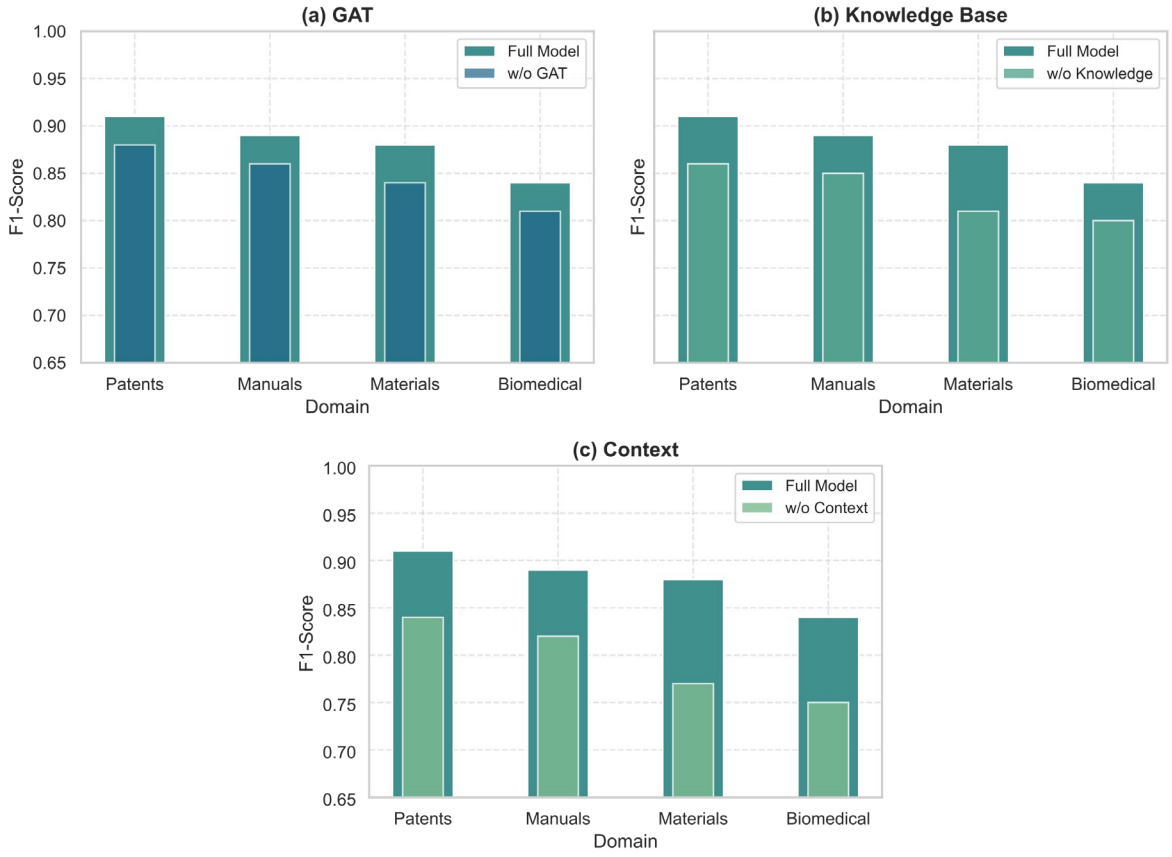

Figure 4. Ablation Study of Model Components: (a) GAT; (b) Knowledge Base; (c) Context.

Figure 5 shows the complete hyperparameter sensitivity analysis, including the aforementioned architecture-level ablation experiments. As shown in Figure 5(a), multi-head attention can increase the number of heads in the model by enhancing the model's expressiveness and stability. However, too many heads may lead to diminishing returns and cause overfitting. As shown in Figure 5(b), choosing the ideal neighbor size directly affects the trade-off between expressiveness and computational cost. If the neighborhood is too large, local

https://doi.org/10.64972/jiic.2025v3.219p6s:69-82

minima will appear; if the neighborhood is too large, the node signal will weaken. As shown in Figure 5(c), the relationship between the learning rate and convergence indicates that a longer step size can accelerate optimization without making the network unstable; conversely, the convergence process is either too slow or ineffective.

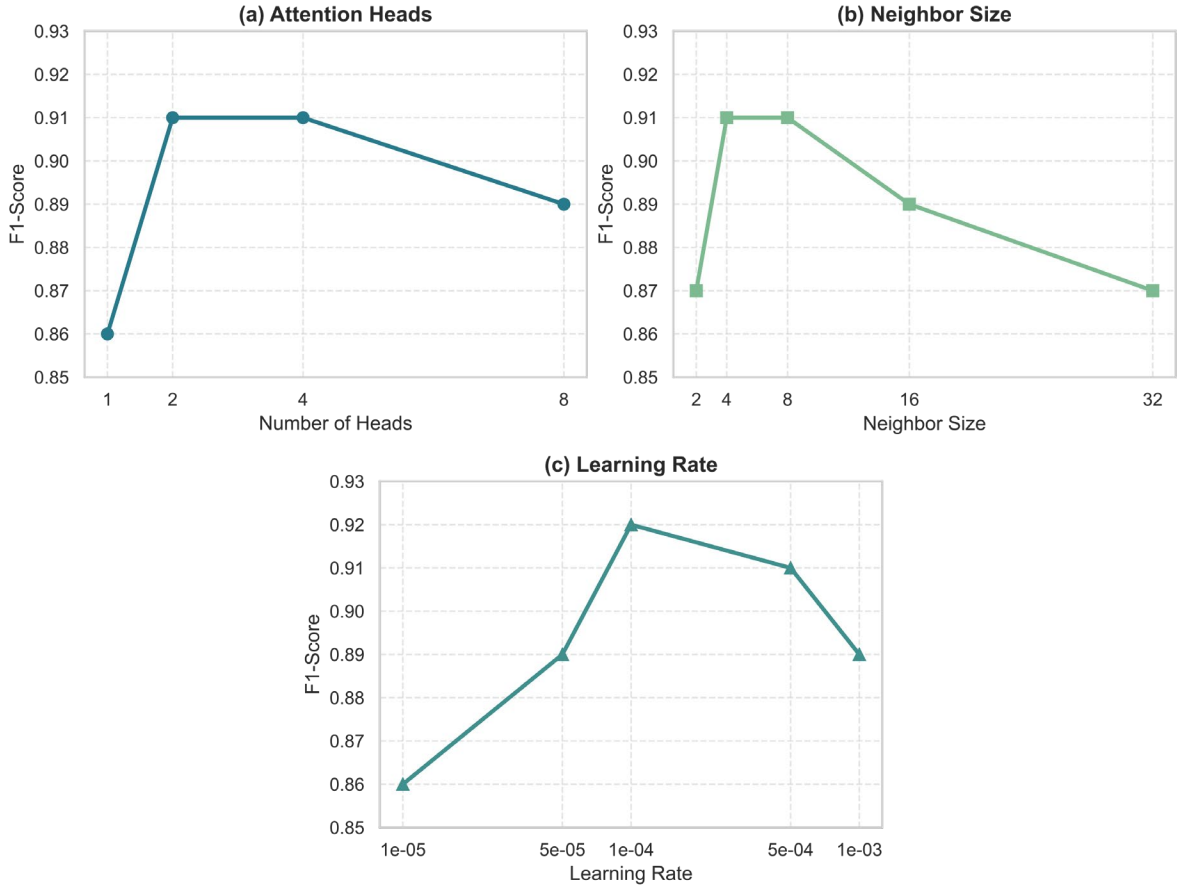

Figure 5. Hyperparameter Sensitivity Analysis: (a) Attention Heads; (b) Neighbor Size; (c) Learning Rate.

According to the above results, and since each ablation set supports at least five independent seed experiments, all modules and parameters are essential components of the entire system on the robust datasets across all six graphs. Deterioration indicates that using graph-based relational reasoning or external knowledge and background modeling is not necessary unless it is very obvious. Hyperparameter analysis indicates that to maintain the model's optimal performance in entity linking across different domains, some adjustments to the default values are necessary.

Case Studies and Limitations

How the model's flexible entity linking is used in various technical fields has been specifically applied. As shown in Figure 6(a), the accuracy of semiconductor patents is 93.8%, successfully addressing the complex dependencies and dense terminology issues of hierarchical entity references. As shown in Figure 6(b), 92% of these cases were correctly resolved, indicating the robustness of normalization to domain changes. Figure 6(c) shows the entity linking of phrases in the emerging field of advanced materials. For example, "graphene-reinforced composite interface" does not exist in traditional ontologies, but it has been accurately mapped to the ontology with an accuracy of over 90%, indicating that this structure can be generalized through local semantic combinations.

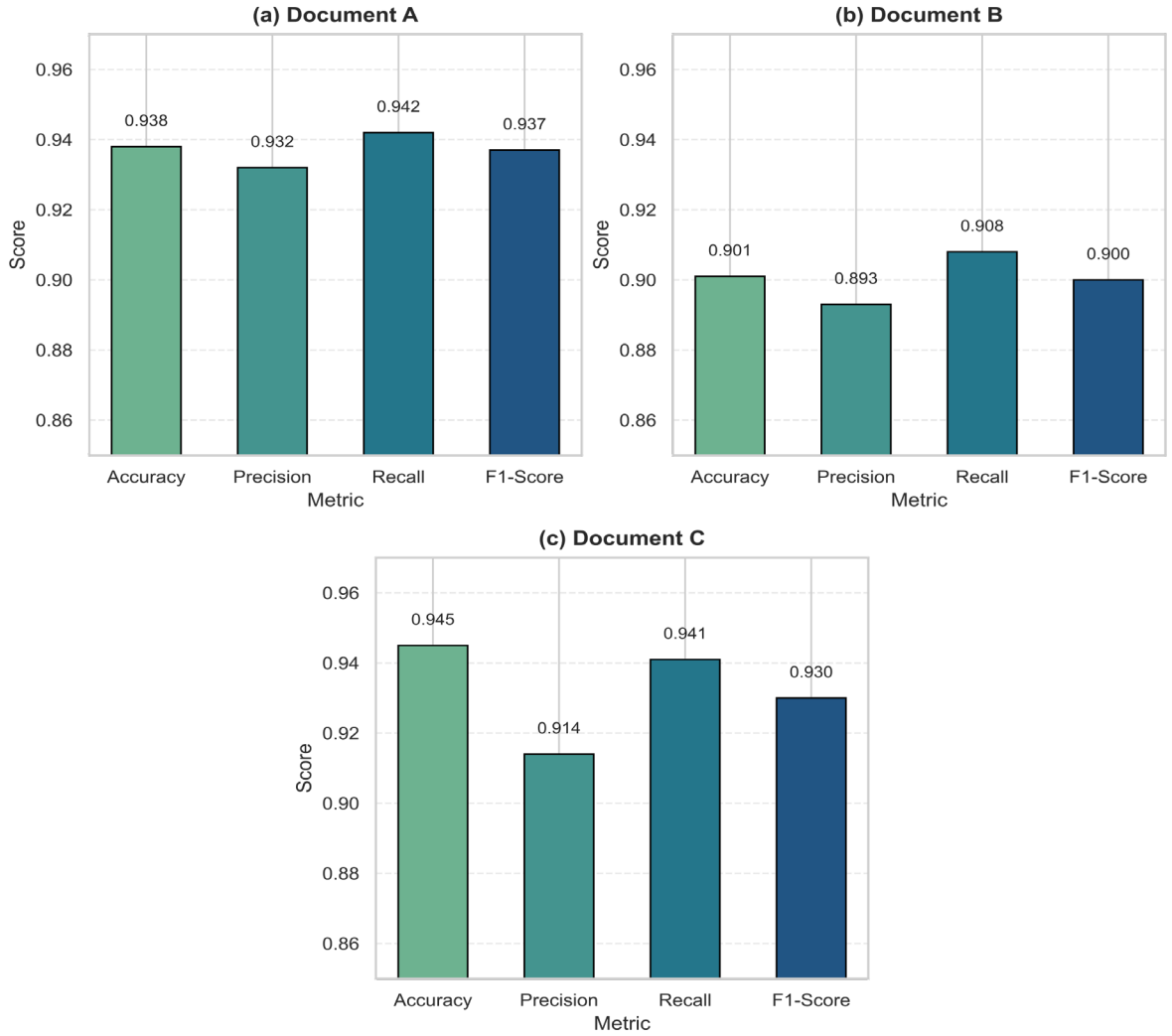

Figure 6. Case Study Visualizations: (a) Document A; (b) Document B; (c) Document C.

Biomedical and legal materials have these defects. As shown in Figure 7(a), the accuracy of the domain-specific modules is 85.6% and 81.3%, respectively, with the most common errors occurring in sections with ambiguous abbreviations (e.g., "ACE," which can have multiple interpretations depending on the context but lacks sufficient co-reference signals). As shown in Figure 7(b), the false negative rate for compressed scientific abstracts rises to 22%. In these short, complex texts, entity definitions are difficult to find, as they may be hidden or vary between different sections.

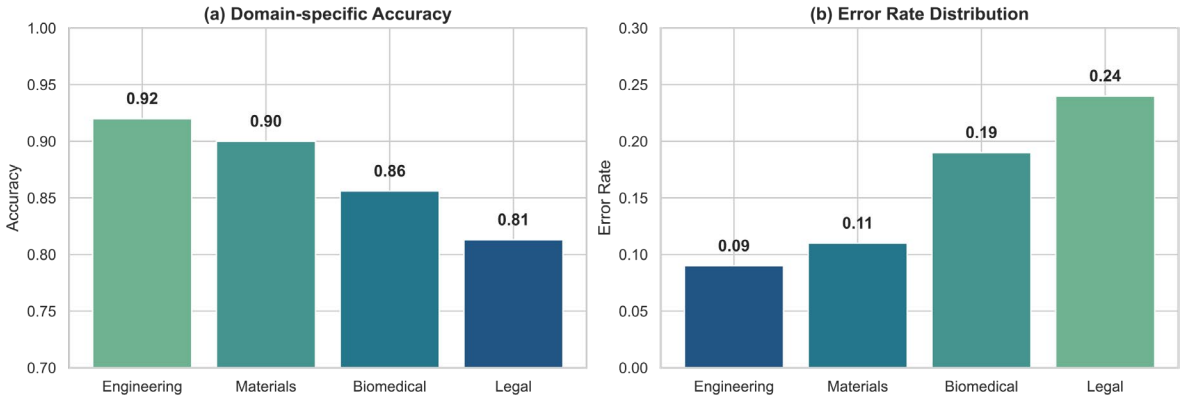

Figure 7. Cross-domain Analysis: (a) Domain-specific Accuracy; (b) Error Rate Distribution.

The aforementioned results stem from several systemic issues, including non-standard abbreviations, unclear cross-references within consistency documents, and occasional inconsistencies in actual annotations. The accuracy of high-performance engineering and industrial textbooks has consistently exceeded 90%, and they are considered enterprise-ready. However, less formal and rapidly changing technical drafts require feedback-driven knowledge integration.

According to the above analysis, many issues fall into one of the following three categories: high complexity of the field, frequent changes in terminology, or lack of background knowledge. Documents generated from preliminary research or cross-organizational collaboration may include non-standard, emerging entities more quickly than organized knowledge resources. In this case, manual review often uses incomplete information for linking, and occasionally, unusual errors occur, such as the most recent match not being the expected reference. The returned entity links, after post-processing constraints or professional review, will have an accuracy level that is acceptable for operational purposes. When the domain ontology and contextual signals are clearly defined, the maximum effect will be achieved. In order to ensure the reliability of the system at all stages of industrial and scientific applications, further research is still needed in the areas of discourse interpretation and continuous learning.

Conclusion

This paper introduces a new framework for graph-based technical entity linking, model architecture, and empirical data. This method addresses the limitations of traditional pipeline methods by systematically integrating various graph attention mechanisms and deep context fusion techniques. It has improved the ability to analyze ambiguous and overlapping entities and has been widely applied in engineering and scientific fields. The new model still outperforms the existing baseline model in terms of accuracy, precision, and recall. This is especially applicable in cases where a large number of technical terms, uncommon expressions, or knowledge resources are used. Further ablation and analysis of the system's sensitivity were conducted to confirm the necessity of adaptive graph reasoning and domain-aware contextual modeling. Large-scale cross-domain experiments further demonstrate the feasibility and adaptability of this method in industrial and scientific fields.

Theoretical support provides a new systematic framework for integrating structural and semantic evidence. Relational perception propagation in knowledge-enhanced graphs helps identify potential dependencies within complex technical corpora and dynamically refines contextual understanding at both document and local levels. The two will collaborate to ensure the system maintains high reliability, even as the vocabulary of new research and industrial applications rapidly changes. The generalization ability of link logic is supported by interpretable attention mechanisms and detailed ablation studies, making it a reliable tool for expert involvement in applications and automated document processing.

The results of this study provide some research directions for the future. Strengthening discourse perception and temporal entity modeling to ensure stable connections between entities across different versions of documents and multiple collaborators. Real-time knowledge expansion and incremental ontology learning can be combined to improve the performance of the ontology in a constantly changing technological environment. To improve efficiency, a human-machine collaborative review system can be used to manage the connections between new entities and advanced entities. Graph-based entity linking will continue to benefit fields such as science and technology, engineering innovation, and enterprise knowledge management.

Author Contributions

Marcel Barczyk contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, project administration, and funding acquisition. Edmund Kapuściński contributes to conceptualization, methodology, software, validation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Najafabadi, M. K., Chen, R. A., Rezazadeh, J., Beheshti, A., & Shabani, N. (2025). From theory to practice: The evolution and comparative analysis of homogeneous vs. heterogeneous Graph Neural Networks in recommender systems. Neurocomputing, 624, 129446. https://doi.org/10.1016/j.neucom.2025.129446

[2] Ye, Z., Kumar, Y. J., Sing, G. O., Song, F., & Wang, J. (2022). A comprehensive survey of graph neural networks for knowledge graphs. Ieee Access, 10, 75729-75741. https://doi.org/10.1109/ACCESS.2022.3191784

[3] Shen, W., Li, Y., Liu, Y., Han, J., Wang, J., & Yuan, X. (2021). Entity linking meets deep learning: Techniques and solutions. IEEE Transactions on Knowledge and Data Engineering, 35(3), 2556-2578. https://doi.org/10.1109/TKDE.2021.3117715

[4] Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. (2023). A comprehensive survey on automatic knowledge graph construction. ACM Computing Surveys, 56(4), 1-62. https://doi.org/10.1145/3618295

[5] Abu-Salih, B., & Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. Heliyon, 10(3). https://doi.org/10.1016/j.heliyon.2024.e25383 External Link

[6] Wen, Z., & Fang, Y. (2023, July). Augmenting low-resource text classification with graph-grounded pre-training and prompting. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 506-516). https://doi.org/10.1145/3539618.3591641

[7] Wang, B., Tang, D., & Wu, Q. (2022, July). Improved Pre-training and Semi-supervised Learning for Domain-specific Chinese Named Entity Recognition. In 2022 4th International Conference on Applied Machine Learning (ICAML) (pp. 142-145). IEEE. https://doi.org/10.1109/ICAML57167.2022.00034

[8] Wang, P., Zhu, Z., Chen, Q., & Dai, W. (2024). Text reasoning chain extraction for multi-hop question answering. Tsinghua Science and Technology, 29(4), 959-970. https://doi.org/10.26599/TST.2023.9010060

[9] Cocchieri, A., Frisoni, G., Galindo, M. M., Moro, G., Tagliavini, G., & Candoli, F. (2025, April). Openbioner: Lightweight open-domain biomedical named entity recognition through entity type description. In Findings of the Association for Computational Linguistics: NAACL 2025 (pp. 818-837). https://doi.org/10.18653/v1/2025.findings-naacl.47

[10] Zhao, Y., Zhou, H., Zhang, A., Xie, R., Li, Q., & Zhuang, F. (2022). Connecting embeddings based on multiplex relational graph attention networks for knowledge graph entity typing. IEEE Transactions on Knowledge and Data Engineering, 35(5), 4608-4620. https://doi.org/10.1109/TKDE.2022.3142056

[11] Wang, X., Chen, L., Zhu, W., Ni, Y., Xie, G., Yang, D., & Xiao, Y. (2023). Multi-task entity linking with supervision from a taxonomy. Knowledge and Information Systems, 65(10), 4335-4358. https://doi.org/10.1007/s10115-023-01905-7

[12] Rezaei, M., & Khosravi, A. (2021). Robust Entity Linking and Disambiguation in Noisy, Automatically Extracted Knowledge Graphs. Transactions on Embedded Systems, Real-Time Computing, and Applications, 11(8), 16-27. https://doi.org/10.1007/978-981-97-8620-6_13

[13] Bouarroudj, W., Boufaida, Z., & Bellatreche, L. (2022). Named entity disambiguation in short texts over knowledge graphs. Knowledge and Information Systems, 64(2), 325-351. https://doi.org/10.1007/s10115-021-01642-9

[14] Zhao, Y., Xue, S., Li, X., & Duan, H. (2025). Multi-semantic fusion of heterogeneous graph neural network. Applied Soft Computing, 114402. https://doi.org/10.1016/j.asoc.2025.114402

[15] Möller, C., Lehmann, J., & Usbeck, R. (2022). Survey on english entity linking on wikidata: Datasets and approaches. Semantic Web, 13(6), 925-966. https://doi.org/10.3233/SW-212865

[16] Jiang, X., Xu, C., Shen, Y., Wang, Y., Su, F., Shi, Z., ... & Shen, H. (2024, May). Toward practical entity alignment method design: Insights from new highly heterogeneous knowledge graph datasets. In Proceedings of the ACM Web Conference 2024 (pp. 2325-2336). https://doi.org/10.1145/3589334.3645720

[17] Santini, C., Gesese, G. A., Peroni, S., Gangemi, A., Sack, H., & Alam, M. (2022). A knowledge graph embeddings based approach for author name disambiguation using literals. Scientometrics, 127(8), 4887-4912. https://doi.org/10.1007/s11192-022-04426-2

[18] Shi, J., Yuan, Z., Guo, W., Ma, C., Chen, J., & Zhang, M. (2023). Knowledge-graph-enabled biomedical entity linking: a survey. World Wide Web, 26(5), 2593-2622. https://doi.org/10.1007/s11280-023-01144-4

[19] Pham, P., Nguyen, L. T., Pedrycz, W., & Vo, B. (2023). Deep learning, graph-based text representation and classification: a survey, perspectives and challenges. Artificial Intelligence Review, 56(6), 4893-4927. https://doi.org/10.1007/s10462-022-10265-7

[20] Zhong, Z., Barkova, A., & Mottin, D. (2025). Knowledge-augmented graph machine learning for drug discovery: a survey. ACM Computing Surveys, 57(12), 1-38. https://doi.org/10.1145/3744237

[21] Du, K., Yang, B., Wang, S., Chang, Y., Li, S., & Yi, G. (2022). Relation extraction for manufacturing knowledge graphs based on feature fusion of attention mechanism and graph convolution network. Knowledge-Based Systems, 255, 109703. https://doi.org/10.1016/j.knosys.2022.109703

[22] Zhao, Y., Dong, J., Wang, W., & Duan, H. (2025). A multi-typed multi-relational heterogeneous graph neural network model for complex networks. Knowledge-Based Systems, 114291. https://doi.org/10.1016/j.knosys.2025.114291

[23] Nam, D., Kim, J., Yoon, J., Song, C., Kim, S., & Song, M. (2024). Examining knowledge entities and its relationships based on citation sentences using a multi-anchor bipartite network. Scientometrics, 129(11), 7197-7228. https://doi.org/10.1007/s11192-023-04824-0

[24] Wang, Y. C., Chuang, C. M., Wu, C. K., Pan, C. L., & Tsai, R. T. H. (2022). Cross-language article linking with deep neural network based paragraph encoding. Computer Speech & Language, 72, 101279. https://doi.org/10.1016/j.csl.2021.101279

[25] Boussouf, M., & Chafik, H. (2021). Optimizing Query Processing in Large-Scale Graph-Based Knowledge Bases Using Advanced Traversal Techniques. Journal of Data Mining, Knowledge Discovery, and Decision Support Systems, 11(1), 15-29. https://doi.org/10.1142/S2972370124300024

[26] Dai, Q., Zhong, J., Li, K., Li, R., Wang, C., Lv, L., ... & Li, X. (2024, December). Contrastive learning enhanced graph relation representation for document-level relation extraction. In 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 4885-4892). IEEE. https://doi.org/10.1109/BIBM62325.2024.10822629

[27] Fu, T., & Zhou, G. (2025). Temporal knowledge completion enhanced self-supervised entity alignment. Journal of Intelligent Information Systems, 63(1), 43-62. https://doi.org/10.1007/s10844-024-00878-5

[28] Xie, C., Deng, L., Tang, Z., & He, J. (2024, December). Fusion and Construction Strategy of Knowledge Graphs from Multi-Source Data. In 2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-7). IEEE. https://doi.org/10.1109/ICMNWC63764.2024.10872219

[29] Sun, X., Liang, K., Zhou, S., & Chen, J. (2025, November). Topic-Enhanced Instruction Tuning for Automatic Knowledge Graph Construction. In 2025 IEEE International Conference on Knowledge Graph (ICKG) (pp. 348-354). IEEE. https://doi.org/10.1109/ICKG66886.2025.00052

[30] Phung, D., Nguyen, T. N., & Nguyen, T. H. (2021, June). Hierarchical graph convolutional networks for jointly resolving cross-document coreference of entity and event mentions. In Proceedings of the Fifteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-15) (pp. 32-41). https://doi.org/10.18653/v1/2021.textgraphs-1.4

[31] Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE transactions on neural networks and learning systems, 33(2), 494-514. https://doi.org/10.1109/TNNLS.2021.3070843