Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# FCA-Attention: Neuro-Symbolic Integration via Concept Lattice Priors for Joint Clinical Entity and Relation Extraction

**Chunlei Cheng**
Jiangxi University of Traditional Chinese Medicine
**Shujun Tan**
2958786663@qq.com
Jiangxi University of Traditional Chinese Medicine
**Xinyi Xiong**
Jiangxi University of Traditional Chinese Medicine
**Cong Cheng**
Jiangxi University of Traditional Chinese Medicine

**Research Article**

**Keywords:** Joint Entity and Relation Extraction, Formal Concept Analysis, Knowledge and Data Fusion, Attention Mechanism, Neuro-Symbolic Integration, Medical Information Extraction

**Posted Date:** September 9th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10833880/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# FCA-Attention: Neuro-Symbolic Integration via Concept Lattice Priors for Joint Clinical Entity and Relation Extraction

Chunlei Cheng[1,2] ,Shujun Tan[1*] ,Xinyi Xiong[1] ,Cong Cheng[1]

[1]School of Intelligent Medicine and Information Engineering, Jiangxi University of Chinese Medicine, Nanchang 330004, China

[2]Jiangxi Provincial Key Laboratory of Artificial Intelligence in Traditional Chinese Medicine, Nanchang 330004, China

* Corresponding author:Shujun Tan. E-mail address: 2958786663@qq.com

**Abstract:** Joint entity and relation extraction from medical texts is essential for constructing medical knowledge systems. However, purely data-driven models often struggle with complex terminology, implicit relations, limited annotations, and insufficient interpretability. To address these limitations, we propose FCA-Attention, a relation-aware neuro-symbolic framework that integrates formal concept analysis with contextual language representations. The framework constructs a formal concept repository from relation triples in the training corpus, capturing associations between entity extents and relation intents and organizing their hierarchical dependencies through a concept lattice. Its partial-order and hierarchical features are encoded as continuous vectors and fused with BERT representations through concept-aware attention, enabling joint entity recognition and relation extraction within a unified architecture. Unlike methods that rely on external knowledge graphs, FCA-Attention derives structural priors directly from the task data. Experiments on the CBLUE medical benchmark show absolute F1 improvements of 0.74% and 1.91% for entity recognition and relation extraction, respectively. Under low-resource settings using less than 10% of the training data, the gains are more pronounced, indicating improved data utilization. Ablation studies and case analyses further examine how concept scale and quality affect performance. On the general-domain DuIE benchmark, FCA-Attention achieves an absolute F1 improvement of 1.13% in relation extraction, supporting its cross-domain applicability. These results show that concept lattice priors can reduce the representational gap between neural and symbolic models without additional external knowledge, providing an interpretable and reproducible approach to medical information extraction.

**Keywords:** Joint Entity and Relation Extraction, Formal Concept Analysis, Knowledge and Data Fusion,Attention Mechanism,Neuro-Symbolic Integration,Medical Information Extraction

## 1. Introduction

Medical text information extraction serves as a foundational task for the structuration of medical knowledge. The automatic extraction of "entity-relation-entity" triplets from unstructured clinical texts provides crucial data support for higher-level applications, including knowledge graph construction, intelligent question answering, and clinical decision support[1]. As healthcare informatization accelerates, text data such as electronic medical records(EMRs) and clinical literature are increasing exponentially, containing massive amounts of diagnosis and treatment knowledge as well as clinical experience. Thus, how to efficiently and accurately perform domain-specific triplet extraction possesses vital theoretical and practical value[2]. However, joint entity and relation extraction from medical texts faces a series of unique challenges. First, the medical terminology system is inherently complex, with prominent issues of ambiguity and synonymy[3]. A single medical concept often has multiple surface expressions. For instance, in Chinese medical texts, "myocardial infarction" (心肌梗死) is frequently abbreviated as "Xin Geng" (心梗). The flexible nature of such medical entity mentions exacerbates the challenge of entity recognition[4]. Second, relation expressions are implicit and highly logical. Relations in medical texts are often embedded within the narrative logic[5]. For example, symptoms such as "fever, cough, and elevated white blood cells" may collectively indicate a "pulmonary infection," requiring the model to possess capabilities for semantic composition and reasoning. Furthermore, the phenomenon of overlapping relations is widespread; a single entity may participate in multiple relations, which poses a significant challenge to the design of encoding and decoding mechanisms[6]. In addition, there exists a strong semantic coupling between entities and relations[7]. Within the medical knowledge framework, entity types naturally constrain relation types. For instance, only a "Disease-Symptom" entity pair can constitute a "Clinical Manifestation" relation. However, traditional pipeline methods decouple the two tasks, leading to irreversible error propagation[8]. Although existing joint models attempt parameter sharing, they are mostly confined to shallow feature sharing and fail to deeply integrate domain knowledge into the learning process[9]. Therefore, fully incorporating domain knowledge holds significant value in the field of medical text processing.

At present, pre-trained language models(PLMs) represented by BERT have become the mainstream paradigm in natural language processing due to their powerful contextual representation capabilities[10]. However, in the medical domain, the cost of acquiring high-quality annotated data is prohibitively high. Purely data-driven pre-trained models struggle to learn deep semantic patterns solely from limited data, which restricts their extraction performance, generalization capability, and interpretability[11]. Formal Concept Analysis(FCA), as a classical knowledge representation theory[12], perfectly compensates for this deficiency in a precise and interpretable manner. Specifically, the extension of a concept provides a basis for instance-level disambiguation, intensional relations define a logical reasoning framework, and the concept lattice structure constrains the valid combinations of entities and relations. Deeply integrating structured domain knowledge with data-driven methods is expected to break through the performance bottlenecks of medical information extraction[13].

Based on the above analysis, this paper proposes a relation-aware attention model incorporating Formal Concept Analysis (FCA-Attention).First, the model constructs a structured and computable relational formal concept repository, where each concept is formally defined by its extension (instance sets) and intension (relation sets).Then, through a concept-aware attention mechanism, the model deeply integrates symbolic conceptual knowledge into the BERT encoding process while preserving BERT's general semantic understanding capabilities, thereby achieving semantic alignment between text representations and the concept space.Finally, entity recognition and relation extraction are jointly learned within a unified framework, aiming to enhance the extraction performance, generalizability, and interpretability of complex semantic relations in medical texts.

The main contributions of this paper are summarized as follows: (1) We propose a novel deep knowledge-data fusion model, termed FCA-Attention, which deeply integrates Formal Concept Analysis with attention mechanisms, thereby providing an efficient new paradigm of heterogeneous information fusion for joint entity and relation extraction in the medical domain.(2) We design a concept-aware attention fusion mechanism that allows structured conceptual knowledge to dynamically enhance textual representations, achieving an adaptive integration of knowledge and data.(3) We conduct comprehensive experiments on both the CBLUE (medical) and DuIE (general) benchmarks, verifying the effectiveness, generalization, and interpretability of the proposed model.

## 2. Related Work

### 2.1 Medical Text Information Extraction

Medical text information extraction primarily consists of two subtasks: Named Entity Recognition(NER) and Relation Extraction(RE). In terms of NER, early research primarily adopted methods combining dictionaries and rules, for instance, utilizing medical dictionaries together with conditional random fields(CRF) to recognize entities such as diseases, symptoms, and drugs[14]. With the advancement of deep learning(DL), methods based on recurrent neural networks(RNN) and their variants have gradually become mainstream[15]. In recent years, PLMs represented by BERT have achieved significant progress in medical NER by virtue of their powerful contextual representation capability[16]. However, medical texts feature a vast and complex terminology system with flexible and diverse expressions, exhibiting phenomena such as polysemy and synonymy[17]. When annotated resources are limited, purely data-driven models still face challenges including entity boundary ambiguity and difficulties in recognizing rare terms[18].

In terms of RE, traditional methods can be divided into two categories: pipeline methods and joint extraction methods. Pipeline methods first identify entities and then classify relations. They are simple to implement but suffer from the error propagation problem[8]. Joint extraction methods attempt to simultaneously perform named NER and RE through parameter sharing or a unified labeling framework, such as TPLinker[19] and PRGC[20]. However, most existing joint models remain largely confined to shallow feature sharing and fail to fully exploit the inherent constraints that entity types impose on relation types[9]. Especially in the medical domain, there exist multi-level conceptual associations among "disease-symptom-drug-examination", making the deep incorporation of

conceptual knowledge even more crucial for joint extraction tasks[21]. How to deeply integrate structured domain knowledge into the joint extraction process while endowing the model with interpretability remains a challenging research issue.

### 2.2 Formal Concept Analysis and Relational Formal Concept

FCA is a mathematical tool based on lattice theory, used to discover conceptual structures from data and construct concept hierarchies[22]. Its core idea originates from the philosophical definition that a "concept" is composed of an "extension" (aobject sets) and an "intension" (attribute sets). Given a formal context $K=(G,M,I)$, where $G$ is object sets, $M$ is attribute sets, and $I \subseteq G \times M$ denotes a binary relation between objects and attributes, FCA can derive formal concepts through a Galois connection. Each formal concept $C=(A,B)$ satisfies $A \subseteq G$, $B \subseteq M$, and $A'=B$ (where $A'$ denotes the attribute sets common to all objects in $A$), $B'=A$ (where $B'$ denotes the object sets that possess all attributes in $B$). All concepts, under the partial order relation, form a complete lattice—the concept lattice(CL)—which clearly reveals the generalization and specialization relationships among concepts.

To accommodate the representation requirements of multi-relational data such as knowledge graphs(KG), the author, in earlier work, extended and proposed the Relational Formal Concept[23][24]. In the context of this study, a Relational Formal Context is defined as $K=(G,M,RI)$, where G is subject entitie sets (e.g., diseases, examinations), $M$ is object entitie sets (e.g., symptoms, drugs), and $RI$ is multi-valued relation sets defined on $G \times M$. From this, two core types of concepts can be derived: the subject concept, defined as $SC=(G_s,R_s)$, where its extension $G_s$is subject entitiey sets sharing the common relational features $R_s$, and its intension $R_s$ is relation sets common to all objects in $G_s$; and the object concept, defined as $OC=(M_o,R_o)$, where its extension $M_o$ is object entitiey sets serving as relational objects, and its intension $R_o$ is likewise a set of relations. Relational formal concepts internalize "relations" as concept intensions, providing a unified formal description framework for both entities and relations.

### 2.3 Attention Mechanism and Knowledge Fusion

Attention mechanisms, especially the self-attention in the Transformer, have become a core technology for processing sequential data in deep learning. By calculating the correlation weights among input elements, it enables the model to dynamically focus on key information and endows it with the capability to model long-range dependencies[25]. However, when processing specialized texts such as medical documents, purely data-driven attention often struggles to focus on truly discriminative semantic units due to limited training data[26].

To incorporate external knowledge, researchers have proposed a variety of knowledge fusion methods[27]. One category of approaches focuses on knowledge enhancement during the pre-training phase. For instance, K-BERT injects knowledge graph triplets into sentences via a visibility matrix[28]; KnowBERT dynamically integrates knowledge into contextual representations using an entity linker and a knowledge attention mechanism[29]; K-Adapter inserts knowledge between Transformer layers through adapter modules while keeping the pre-trained parameters frozen[30]; and ERNIE learns language and knowledge representations through masked entity prediction[31]. However, the knowledge

incorporated by these methods is static and global, making it difficult to perform context-dependent dynamic invocation during inference. Another category of methods introduces knowledge during the downstream task fine-tuning or inference stage. These approaches retrieve information from external knowledge bases and integrate the retrieved results using attention mechanisms, such as RAG[32], REALM[33], and RETRO[34]. However, these methods typically treat knowledge as an independent module, where the interaction between knowledge and text mostly remains at representation concatenation or shallow attention weighting. Consequently, they fail to achieve a deep alignment between knowledge logic and the neural network's computational process.

Knowledge Graph-guided Attention(KGA) is a recently proposed test-time knowledge fusion method[35]. Inspired by the dual-stream theory in cognitive neuroscience, it achieves dynamic knowledge fusion through bidirectional information aggregation without modifying the parameters of large language models(LLMs). Specifically, the outward aggregation enables the input text to query knowledge triplets, while the inward aggregation allows the knowledge triplets to evaluate their relevance to the text and generate adaptive weights, thereby modulating the fusion process. KGA textualizes knowledge and incorporates it into the attention computation in parallel with the input. By reusing the original projection matrices of the LLM, it achieves a closed loop of "textual context-driven knowledge retrieval and knowledge-guided attention focusing." This method has achieved significant improvements in tasks such as knowledge-based question answering and model editing, demonstrating the immense potential of test-time knowledge fusion.

The proposed FCA-Attention in this study conceptually echoes KGA but also presents notable differences. While KGA emphasizes parameter-free fusion at test time, FCA-Attention deeply integrates structured knowledge into the model training process through trainable concept vectors. This makes it more suitable for medical scenarios where domain knowledge is relatively stable and requires long-term iterative optimization. Both approaches collectively point towards the cutting-edge direction of "neuro-symbolic synergy," yet FCA-Attention possesses unique advantages in interpretability and domain adaptability.

## 3. Construction of Medical Relational Formal Context for LLM

The sources of the medical relational formal context may include the following:(1) Structured knowledge bases: For instance, standardized dictionaries like CHIP-CDN can serve as the authoritative extent of disease concepts, along with triplet data derived from medical knowledge graphs, such as (Aspirin, Indication, Fever). (2) Semi-structured standard literature: This includes medical textbooks and clinical guidelines, such as Internal Medicine and Clinical Practice Guidelines, which contain a wealth of medical knowledge definitions and relational descriptions. (3) Unstructured clinical texts: Examples include EMRs and clinical literature. While these texts feature flexible linguistic expressions and closely reflect real-world clinical practice, they contain considerable noise.

The quality and scale of the constructed relational formal context directly impact the model's extraction performance. To objectively validate the effectiveness of the knowledge-data fusion proposed in this paper, the relational formal context in our experiments is constructed directly using the

triplets from the training data. No external knowledge is introduced, thereby avoiding direct interference with the extraction test results.

The objective is to construct a structured and computable relational formal concept knowledge base. This knowledge base must simultaneously support both NER and RE tasks, providing a dynamically queryable conceptual memory resource for the subsequent FCA-Attention model. The construction process consists of three stages: concept definition, concept lattice construction, and concept vectorization.

### 3.1 Formal Definition of Concepts

Let $K=(G, M, RI)$ be a relational formal context, where $G$ is the set of subject entities, $M$ is the set of object entities, and $RI$ denotes the set of relations or relational predicates. Based on this framework, we define two categories of core concepts:

Subject Concept $SC=(G_s, R_s)$: Here, $G_s \subseteq G$ is subject entitiey sets sharing identical relational characteristics, termed the extent of the concept; $R_s \subseteq RI$ is relation sets jointly participated in by these subject entities, termed the intent of the concept. For instance, the extent of the subject concept "Cardiovascular Disease" includes diseases such as "Coronary Heart Disease," "Myocardial Infarction," and "Hypertension," whereas its intent encompasses relations like "Clinical Manifestation (Chest Pain)" and "Complication (Heart Failure)."

Object Concept $OC=(M_o, R_o)$: Here, $M_o \subseteq M$ is entitiey sets serving as relation objects, termed the extent of the concept; $R_o \subseteq RI$ is relation sets participated in by these object entities, termed the intent of the concept. For instance, the extent of the object concept "Anticoagulant Drugs" includes medications such as "Aspirin," "Warfarin," and "Clopidogrel," while its intent encompasses relations like "Indication (Coronary Heart Disease)" and "Contraindication (Gastrointestinal Bleeding)."

Through this definition, concepts directly bridge entity types and relation types: the concept to which an entity belongs is determined by its extent, while the concept's intent explicitly dictates the relation types the entity can participate in. For example, if a certain drug belongs to the "Anticoagulants" concept, the "Indication (Coronary Heart Disease)" relation within its intent can be leveraged by the model. Simultaneously, it constrains the potential therapeutic relations between this drug and subject entities such as "Cardiovascular Disease." This design enables the conceptual repository to simultaneously serve both entity classification and relation classification tasks.

### 3.2 Concept Hierarchy

Based on the defined concept sets, we need to further reveal their hierarchical structure, namely, by constructing a concept lattice. The concept lattice organizes concepts via a partial order relation, reflecting the generalization and specialization relationships among them.

For subject concepts $SC_1=(G_{s1}, R_{s1})$ and $SC_2=(G_{s2}, R_{s2})$, the partial order relation is defined as follows: $SC_1 \leqslant SC_2$ if and only if $G_{s1} \subseteq G_{s2}$ (equivalently, $R_{s2} \subseteq R_{s1}$). Intuitively, a concept with a smaller extent and a richer intent is a specialized concept, situated at the lower levels of the concept lattice. Conversely, a concept with a larger extent and a more abstract intent is a generalized concept, positioned at the upper levels. The partial order relation among object concepts is defined similarly.

Intuitively, a concept with a smaller extent and a richer intent is a specialized concept, situated at the lower levels of the concept lattice. Conversely, a concept with a larger extent and a more abstract intent is a generalized concept, positioned at the upper levels. The partial order relation among object concepts is defined similarly.

The construction steps for the concept lattice are as follows:

(1) Concept Computation: Starting from the relational formal context, compute all formal concepts through FCA algorithms (e.g., the Lattice algorithm).

(2) Hasse Diagram Construction: Compute the covering relations among the concepts to form the Hasse diagram of the concept lattice.

(3) Lattice Pruning and Optimization: Prune and optimize the concept lattice to eliminate redundant concepts, retaining only the core concepts that exhibit significant semantic discriminative power.

The constructed concept lattice not only provides a hierarchical organization for knowledge but also lays the foundation for the subsequent attention mechanism: when the model identifies a specific concept, it can automatically activate its generalized concepts via the concept lattice, thereby achieving the smooth diffusion of knowledge.

### 3.3 Concept Vector Representation

To enable the integration of symbolic concepts into neural network models, each concept must be mapped into a low-dimensional dense vector. This paper adopts a strategy that combines semantic encoding and structural encoding to generate a comprehensive vector representation for the concepts.

**Intent Vector:** Each relation or relational predicate in the concept's intent set $R$ is mapped to a pre-trained vector (e.g., using BERT to encode relation names). Subsequently, the intent vector $v_{int}$ is obtained through attention pooling or average pooling. This vector explicitly encodes the relational semantics carried by the concept.

**Extent Vector:** Obtain the vector representations of all entities in the concept's extent set $E$ (derived by encoding entity names or descriptions using BERT). Subsequently, apply mean pooling to generate the extent prototype vector $v_{ext}$. This vector explicitly reflects the instance features of the concept.

**Structure Vector:** Utilizing the positional information within the concept lattice, the vectors of the concept's ancestor and descendant nodes are aggregated to obtain the structural context vector $v_{struc}$. This vector explicitly encodes the hierarchical relations of the concept.

The final concept vector $v_{concept}$ is obtained by concatenating the above three vectors, followed by a linear transformation:

$$\boldsymbol{v}_{concept} = Linear([\boldsymbol{v}_{int}; \boldsymbol{v}_{ext}; \boldsymbol{v}_{struc}]) \quad (1)$$

All concept vectors constitute a queryable "concept memory bank." In the FCA-Attention model, this memory bank serves as an external knowledge source, interacting dynamically with BERT's hidden representations. Meanwhile, the concept vectors can be further fine-tuned during the end-to-end

training process, achieving the joint optimization of knowledge representations and task objectives.

## 4. Design of the FCA-Attention-Based Joint Extraction Model

### 4.1 Overall Model Architecture

The overall architecture of the FCA-Attention model is illustrated in **Figure 1**. It adopts an encoder-decoder paradigm, utilizing BERT as the shared underlying encoder. The parameters of BERT are frozen to preserve its general text understanding capabilities. On top of the encoder, an entity recognition head and a relation extraction head are connected in parallel.

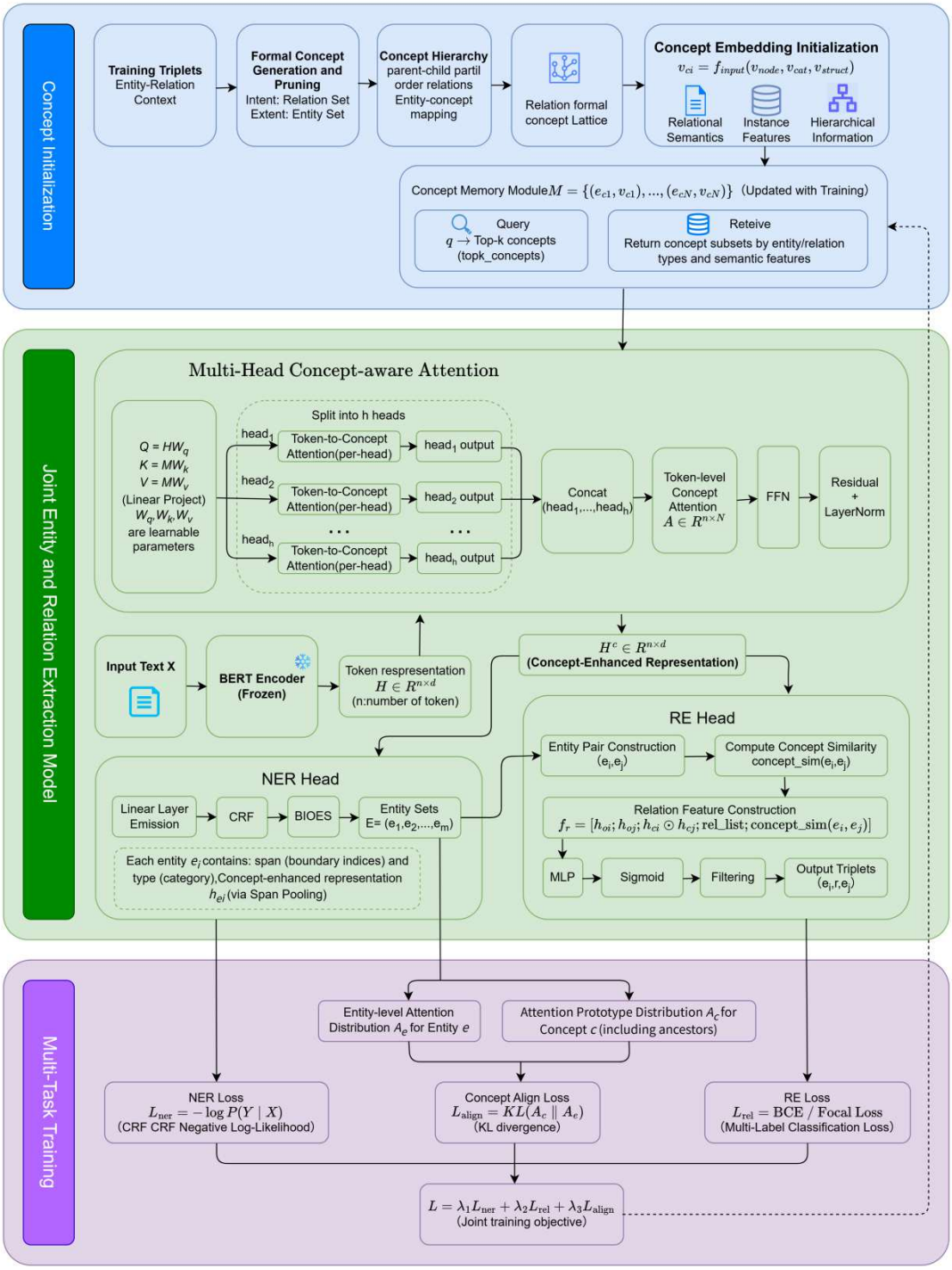


**Figure 1.** Overall architecture of the proposed FCA-Attention model. Solid arrows denote the forward propagation process, while dashed arrows indicate training feedback and gradient backpropagation.

The core innovation lies in the introduction of a concept memory module and a concept-aware attention layer in **Figure 2**. This enables structured conceptual knowledge to dynamically enhance text representations and simultaneously guide the learning of both subtasks.

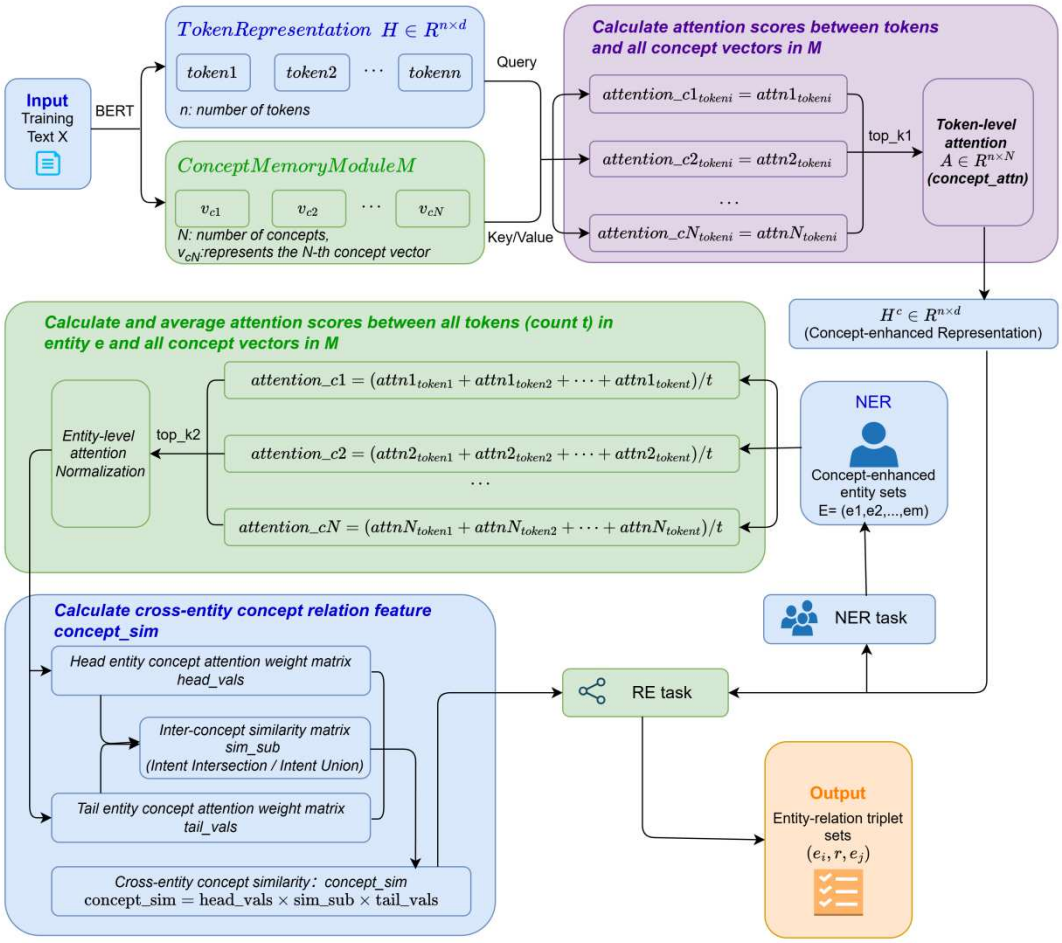



**Figure 2.** Workflow of the concept-aware attention mechanism for joint entity–relation extraction

The workflow of the proposed model proceeds as follows:

(1) Text Encoding: The input text sequence is first encoded by BERT to obtain the context-aware hidden representation matrix $H \in \boldsymbol{R}^{n\times d}$.

(2) Concept Injection: The matrix $H$ is then fed into the concept-aware attention layer. By querying the concept memory module, this layer generates a concept-enhanced representation $H^c \in \boldsymbol{R}^{n\times d}$ for each token.

(3) Joint Decoding and Optimization: $H^c$ is respectively fed into the entity recognition head and the relation extraction head to perform joint decoding. During the training process, the concept memory module is continuously optimized, achieving the co-evolution of knowledge representations and task objectives.

### 4.2 Concept Memory Module

The concept memory module stores the vector representations of all concepts in the relational formal concept repository, denoted as $M = \{v_{c_1}, v_{c_2}, \ldots, v_{c_N}\}$, where $N$ is the total number of concepts, $v_{c_i} \in R^{d_c}$ is the embedding vector of concept $c_i$, and $d_c$ is the dimension of the concept vectors. Each initialized concept vector is generated by fusing the intent vector, extent vector, and structure vector, effectively encoding the concept's relational semantics, instance features, and hierarchical information.

The module supports two basic operations:

**Query:** Given a query vector (typically derived by projecting the textual hidden representations), it calculates the similarity with all concept vectors and returns the *top-k* most relevant concepts (*topk_concepts*) along with their attention weights.

**Retrieval:** Based on the entity type or relation type, it directly returns the corresponding subset of concepts (e.g., all subject concepts or all object concepts) for task-specific attention computation.

### 4.3 Concept-Aware Attention Mechanism

The concept-aware attention layer aims to dynamically integrate symbolic conceptual knowledge into text representations. For the input representation $H$, we first apply linear transformations to obtain the query matrix $Q=HW_q$, the key matrix $K=MW_k$, and the value matrix $V=MW_v$, where $W_q$,$W_k$,$W_v$ are learnable parameters. Then, the cross-attention between the text and the concept space is computed as follows:

$$A = softmax\left(\frac{QK^{\top}}{\sqrt{d_k}}\right), H^c = LayerNorm(H + FFN(AV)) \quad (2)$$

Where $A \in R^{n \times N}$ denotes the attention weight matrix, capturing the semantic correlation between each individual token and concept. $H^c$ signifies the augmented representation infused with conceptual knowledge. Furthermore, $n$ and $N$ indicate the token length of the input text sequence and the total capacity of concepts within the memory module, respectively.

The specific effects of this mechanism on the two subtasks are detailed as follows:

**Guidance on NER**: NER requires accurately determining entity boundaries and their types. In the concept attention weight matrix $A$, if a token is highly relevant to a "disease" concept, the model tends to identify this token as part of a disease entity. Simultaneously, entities within the extent of the same concept should exhibit similarity in the representation space. By sharing concept vectors, the concept attention implicitly constrains the representation consistency of entities of the same category, helping to alleviate the issue of terminology ambiguity.

**Guidance on RE**: The key to RE is identifying whether a predefined relation exists between an entity pair. The intent set of a concept explicitly defines the valid combinations of entity types and relation types. For example, if the intent of the "cardiovascular disease" concept includes "clinical manifestation (chest pain)," then when entities belonging to "cardiovascular disease" and "symptom" co-occur in a sentence, the concept attention will simultaneously enhance the representations of both entities, bringing them closer to the semantic representation of the "clinical manifestation" relation. This guidance enables the model to actively focus on potential relation subjects and objects, reducing

its reliance on explicit predicates.

### 4.4 Joint Decoding Strategy

The model adopts a joint decoding strategy of performing NER prior to RE. However, all steps are entirely based on the shared concept-augmented representation $H^c$, achieving end-to-end training.

(1) **NER Subtask**: The representation $H^c$ is routed into a CRF layer, yielding the optimal label sequence per token based on the BIOES tagging scheme. This label set encompasses all predefined entity categories (such as diseases, symptoms, medications, and medical tests). Subsequently, the decoding stage generates the extracted entity set $E=\{e_1,e_2,...,e_m\}$. Every entity comprises its textual span, semantic type, and a specific concept-augmented representation $h_{ei}$ (derived via pooling the $H^c$ states of all tokens within its span).

(2) **RE Subtask**: For all possible entity pairs $(e_i, e_j)$, the input features for relation classification are constructed as follows:

$$\boldsymbol{f}_{ij} = [h_{e_i}; h_{e_j}; h_{e_i} \odot h_{e_j}; rel_dist; concept_sim(e_i, e_j)] \quad (3)$$

Where $concept_sim\ (e_i,e_j)$ signifies the intent intersection similarity between the concepts to which entities $e_i$ and $e_j$ belong, acting as an inductive bias from prior knowledge. Given that an entity may map to multiple concepts concurrently in our work, the hyperparameter *concept_sim_topk* defines the number of top concepts retrieved for both the head and tail entities during the similarity computation. Subsequently, the formulated feature $f_{ij}$ is passed through a Multi-Layer Perceptron(MLP) with a sigmoid activation to accomplish multi-label classification.

(3) **Final Output:** All relation triplets $(e_i,r,e_j)$ with confidence scores higher than the predefined threshold.

### 4.5 Multi-Task Training and Loss Function

The model adopts a multi-task learning framework, where the total loss function is calculated as the weighted sum of the entity recognition loss, relation extraction loss, and concept alignment loss:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{ner} + \lambda_2 \mathcal{L}_{rel} + \lambda_3 \mathcal{L}_{align} \quad (4)$$

Where $\mathcal{L}_{ner}$, $\mathcal{L}_{rel}$, and $\mathcal{L}_{\text{align}}$ represent the entity recognition loss, relation extraction loss, and concept alignment loss, respectively. Furthermore, $\lambda_1$, $\lambda_2$, and $\lambda_3$ are the weight coefficients for the three types of losses.

**NER Loss $\mathcal{L}_{\text{ner}}$:** The negative log-likelihood loss of the CRF is adopted. It learns the transition constraints between labels by maximizing the conditional probability of the ground-truth label sequence. It is formulated as:

$$\mathcal{L}_{ner} = -logP(Y|X) \quad (5)$$

Where $X$ represents the input text, $Y$ represents the ground-truth BIOES label sequence.

**RE Loss $\mathcal{L}_{rel}$:** The binary cross-entropy (BCE) loss is adopted to perform independent binary classification for all relation types across all entity pairs. Furthermore, Focal Loss is introduced to mitigate the imbalance between positive and negative samples, as well as the uneven distribution of

hard and easy samples. It is formulated as:

$$\mathcal{L}_{rel} = -\sum_{(i,j)} \sum_{r=1}^{|\mathcal{R}|} \alpha_t (1 - p_t)^{\gamma} \log(p_t) \quad (6)$$

Where $(i, j)$ denotes the entity pair, $|\Re|$ represents the number of relation types, and $p_t$ is the model's predicted probability for the ground-truth class. Furthermore, $\alpha_t$ is the class balancing factor, and $\gamma$ is the focusing parameter for hard samples.

**Concept Alignment Loss $\mathcal{L}_{\text{align}}$:** This loss is used to constrain the consistency of concept attention, encouraging entities belonging to the same concept extent to share similar attention patterns. Specifically, for an entity $e$ and its corresponding concept $c$, it calculates and minimizes the $KL$ divergence between the entity-level attention distribution $A_e$ and the attention prototype distribution $A_c$ of concept $c$. It is formulated as:

$$\mathcal{L}_{align} = \frac{1}{|C|} \sum_{e \in C} D_{KL}(A_c \parallel A_e) \quad (7)$$

Through joint training, the model not only learns the mapping from text to entity and relation labels, but the concept memory module also adaptively adjusts itself based on task feedback using the concept alignment constraint. This ultimately achieves the synergistic optimization of symbolic knowledge and neural network parameters.

## 5.Experimental Settings

### 5.1 Datasets and Evaluation Metrics

In this paper, we conduct experiments on CMeIE(Version 2), a core subset of the CBLUE(Chinese Biomedical Language Understanding Evaluation) benchmark dataset, to validate the effectiveness of the proposed method. Furthermore, we perform supplementary experiments on the DuIE(Du Information Extraction) dataset to further verify the model's domain generalization capability.

Specifically, the CMeIE dataset encompasses 44 medical relation categories and 11 entity types, representing the industry's pioneering publicly accessible dataset for Chinese medical relation extraction[36]. Meanwhile, the DuIE dataset comprises 49 general relations and 26 entity types, functioning as a comprehensive dataset for Chinese general extraction[37]. The detailed statistical information of these datasets is summarized in Table 1.

**Table 1**: Statistics of the CMeIE and DuIE Datasets

| Dataset | Training Set | Validation Set | Test Set | Entity / Relation Count | Characteristics |
|---|---|---|---|---|---|
| **CMeIE** | 14,339 | 3,585 | 4,482 | 11/44 | Medical-specific, rigorous annotation, medical knowledge |
| **DuIE** | 171,135 | 20,652 | 101,239 | 26/49 | Large-scale, open-domain, general knowledge |

Since the official test set lacks ground-truth annotations, we re-split the dataset to ensure the

completeness of the evaluation metric calculations. For the CMeIE dataset, we proportionally sampled 12,905 and 1,434 instances from the original training set at a 9:1 ratio to serve as the training and validation sets for our experiments, respectively, while utilizing the original validation set as our test set. Since DuIE is a large-scale dataset, to ensure scale consistency, we evenly sampled 20,000 and 2,000 instances from its original training set to act as our training and test sets, and selected 2,000 instances from its original validation set as our validation set.

We employ Precision (P), Recall (R), and F1-score (F1) as the evaluation metrics. For the entity recognition task, a strict matching criterion is adopted, where a prediction is considered correct only if both its entity boundary and type are perfectly matched. For the relation extraction task, an exact triplet matching criterion is applied, meaning a prediction is deemed correct only when all three components (head entity, relation type, and tail entity) are entirely accurate.

### 5.2 Baseline Models and Experimental Setup

We select the following two categories of baseline models for comparison:

The first category is data-driven models, which include: **Pipeline** (BERT-CRF + Relation Classifier), a pipeline method that first uses BERT-CRF for entity recognition and then conducts relation classification; **TPLinker**[19], a single-stage joint extraction model based on token-pair linking; **PRGC**[20], a joint extraction model based on potential relations and global correspondence mechanisms.

The second category involves knowledge-enhanced models: **KnowBERT**[29], an entity-level knowledge injection model utilizing entity linking and word-to-entity attention for knowledge fusion; **K-Adapter**[30], a parameter-injection model that stores knowledge via supplementary Adapter layers plugged into the pre-trained model; **K-BERT**[28], an input-injection model that introduces knowledge graph triplets into the input and restricts attention with a visibility matrix.

The experimental settings are as follows:

We adopt RoBERTa-zh-Large (jointly released by HIT and iFLYTEK) as our base encoder. The dimensionality of the concept vectors is set to 1024, identical to the hidden size of BERT. We employ the AdamW optimizer, configuring the learning rate at 2e-5 for the BERT layer and 1e-3 for the concept vector layer. During training, the maximum sequence length is limited to 128, and the batch size is set to 32. The model is trained for a maximum of 100 epochs, utilizing an early stopping mechanism that terminates training if the validation F1-score fails to improve for 10 consecutive epochs. Finally, the loss weights are empirically set as $\lambda_1$=1.0, $\lambda_2$=6.0, $\lambda_3$=0.01.

## 6.Experiments and Analysis

### 6.1 Main Results

To verify the effectiveness and generalization capability of the proposed FCA-Attention model in named entity recognition and relation extraction tasks, we conduct comparative experiments on the publicly available medical domain dataset CMeIE and the general domain dataset DuIE. By comparing with existing mainstream baseline models, we systematically evaluate the model's information extraction performance. The experimental results are shown in Table 2 and Table 3.

**Table 2:** Results of Relation Extraction Models on CMeIE

| Model | NER | | | RE | | |
|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 |
| Pipeline | 73.81±0.90 | 71.09±0.92 | 72.42±0.15 | 30.97±0.70 | **44.33±1.26** | 36.45±0.31 |
| PRGC | 58.31±0.37 | 45.84±0.40 | 51.33±0.27 | 31.35±1.31 | 25.80±0.21 | 28.30±0.67 |
| TPLinker | **82.17±0.73** | 30.18±1.04 | 44.14±1.07 | **59.95±1.61** | 19.65±0.51 | 29.59±0.41 |
| FCA-Attention | 73.85±0.28 | **72.85±0.46** | **73.34±0.10** | 38.50±2.07 | 41.12±1.52 | **39.71±0.61** |
| KnowBERT | 71.96±0.95 | 65.03±0.98 | 68.31±0.15 | 32.94±0.78 | 33.17±0.41 | 33.05±0.37 |
| K-Adapter | 72.93±0.64 | 71.90±0.54 | 72.41±0.20 | 36.22±1.33 | 38.66±0.66 | 37.38±0.52 |
| K-BERT | 73.12±0.58 | 70.75±0.78 | 71.91±0.23 | 36.19±1.12 | 36.38±0.86 | 36.26±0.15 |

**Table 3:** Results of Relation Extraction Models on DuIE

| Model | NER | | | RE | | |
|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 |
| Pipeline | 73.17±1.12 | 73.87±1.21 | 73.50±0.25 | 52.64±1.71 | **54.15±1.08** | 53.36±0.37 |
| PRGC | 65.63±1.45 | 59.33±1.86 | 62.29±0.42 | 43.01±2.44 | 42.13±1.53 | 42.52±0.96 |
| TPLinker | **81.39±0.91** | 35.89±1.88 | 49.80±1.99 | **66.95±2.26** | 29.39±1.56 | 40.84±1.92 |
| FCA-Attention | 72.30±0.65 | **75.03±0.92** | **73.64±0.12** | 58.27±0.53 | 51.61±0.48 | **54.74±0.48** |
| KnowBERT | 71.45±1.33 | 41.88±0.95 | 52.80±1.02 | 48.18±2.29 | 19.97±1.03 | 28.23±1.35 |
| K-Adapter | 71.56±1.45 | 74.13±1.20 | 72.80±0.22 | 47.74±2.50 | 51.51±1.15 | 49.51±0.83 |
| K-BERT | 71.96±0.95 | 65.00±1.57 | 67.85±0.83 | 46.38±1.14 | 31.01±2.50 | 37.12±1.82 |

The original frameworks of the TPLinker and PRGC models are primarily designed for joint entity-relation extraction and do not output entity category information. Therefore, for the NER task, we only evaluate these two models based on entity boundary matching. Consequently, their evaluation metrics are slightly higher than those evaluated under the strict matching standard.

The experimental results demonstrate that the proposed FCA-Attention model achieves the highest F1-score on both the CMeIE medical dataset and the DuIE general dataset. This indicates that the introduction of the concept memory module and the concept-aware attention mechanism can effectively enhance the model's ability to model entity and relation semantics, achieving excellent performance in the medical domain while demonstrating strong cross-domain generalization capability.

In our experiments, the model's RE metrics on the CMeIE dataset are relatively low, which is primarily attributed to the highly challenging inherent characteristics of this medical text dataset. Specifically: First, a single sentence typically contains multiple relation triplets, making entity overlapping extremely common. Second, medical sentences are generally lengthy with complex modifiers and are filled with numerous nested and coordinate entities, resulting in highly ambiguous entity boundaries. Third, the relation expressions between entities are often highly implicit. Finally, the relation categories exhibit an extreme long-tail distribution, and some categories are highly semantically similar, which further increases the difficulty of classification and extraction for the

model.

### 6.2 Concept Enhancement Experiments

The relation modeling paradigms of PRGC and TPLinker fundamentally differ from our proposed architecture. Specifically, PRGC relies heavily on relation judgment, relation-specific sequence labeling, and a global correspondence matrix for extraction. Conversely, TPLinker unifies entity and relation extraction into an upper-triangular sequence labeling task via a handshaking tagging scheme. Since neither approach explicitly constructs relation feature vectors for entity pairs, the relation concept similarity features cannot be directly transferred. Forcibly incorporating them would necessitate redesigning their relation classification modules, potentially disrupting the core architectures of the original models. Furthermore, implementing the concept alignment loss demands additional entity-concept mapping supervision. This would introduce new training objectives, thereby compromising the fairness of the comparative experiments.

Therefore,to demonstrate the generalizability of external concept knowledge across various RE frameworks, we propose a lightweight enhancement strategy. This strategy injects only the concept memory module, concept-aware attention, and concept memory parameterization into the baselines, reserving the relation concept similarity feature and concept alignment loss as specific enhancements for our main model. The performance of these enhanced models is then compared against the vanilla baselines. Tables 4 and 5 summarize the experimental results.

**Table 4:** Results of Concept-Enhanced Data-Driven Models on CMeIE

| Model | | NER | | | RE | | |
|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 |
| Pipeline | Baseline | **73.81±0.90** | 71.09±0.92 | 72.42±0.15 | 30.97±0.70 | **44.33±1.26** | 36.45±0.31 |
| | Concept-Aug | 73.75±0.89 | **72.06±0.90** | **72.89±0.08** | **33.13±0.66** | 43.75±0.14 | **37.70±0.39** |
| PRGC | Baseline | 58.31±0.37 | 45.84±0.40 | 51.33±0.27 | 31.35±1.31 | 25.80±0.21 | 28.30±0.67 |
| | Concept-Aug | **58.89±0.64** | **49.84±0.68** | **53.99±0.60** | **32.92±1.40** | **29.43±0.31** | **31.06±0.45** |
| TPLinker | Baseline | **82.17±0.73** | 30.18±1.04 | 44.14±1.07 | **59.95±1.61** | 19.65±0.51 | 29.59±0.41 |
| | Concept-Aug | 82.02±1.61 | **35.04±2.48** | **49.05±2.13** | 59.73±2.35 | **23.51±1.82** | **33.68±1.51** |

**Table 5:** Results of Concept-Enhanced Data-Driven Models on DuIE

| Model | | NER | | | RE | | |
|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 |
| Pipeline | Baseline | **73.17±1.12** | 73.87±1.21 | 73.50±0.25 | 52.64±1.71 | **54.15±1.08** | 53.36±0.37 |
| | Concept-Aug | 72.99±1.03 | **74.52±2.13** | **73.73±0.59** | **55.26±1.29** | 53.66±2.01 | **54.41±0.45** |
| PRGC | Baseline | 65.63±1.45 | 59.33±1.86 | 62.29±0.42 | 43.01±2.44 | 42.13±1.53 | 42.52±0.96 |
| | Concept-Aug | **66.63±0.44** | **59.87±0.37** | **63.07±0.36** | **45.27±0.76** | **43.76±0.46** | **44.48±0.61** |
| TPLinker | Baseline | **81.39±0.91** | 35.89±1.88 | 49.80±1.99 | **66.95±2.26** | 29.39±1.56 | 40.84±1.92 |
| | Concept-Aug | 77.98±3.32 | **41.65±1.41** | **54.26±1.20** | 61.11±5.35 | **34.22±1.29** | **43.82±2.98** |

As shown in Tables 4 and 5, after introducing concept knowledge, the overall performance of different models on both the CMeIE medical dataset and the DuIE general-domain dataset exhibits an upward trend. However, the magnitude of improvement and the specific impact patterns vary.

In the NER task, the Pipeline, PRGC, and TPLinker baselines achieve absolute F1-score gains of 0.57%, 2.66%, and 4.91% on the CMeIE dataset, alongside improvements of 0.23%, 0.78%, and 4.46% on the DuIE dataset. Regarding the RE task, their F1-scores are boosted by 1.25%, 2.76%, and 4.09% on CMeIE, and 1.05%, 1.96%, and 2.98% on DuIE, respectively. Across both tasks, the F1 improvements are predominantly driven by gains in Recall. This demonstrates that incorporating concept knowledge empowers the models to uncover more latent entities and relations, thereby mitigating the false-negative issues stemming from conservative prediction behaviors.

The three models show greater improvements in the CMeIE medical domain than in the DuIE general domain, indicating that concept knowledge plays a more significant auxiliary role in knowledge-intensive domains.

In summary, the comprehensive results across both datasets demonstrate that concept knowledge serves as an effective and universal auxiliary component for diverse relation extraction architectures. However, its underlying impact mechanisms vary according to specific model structures and dataset characteristics. For the Pipeline model, conceptual information primarily enriches entity semantic representations, thereby mitigating the well-known error propagation problem from the NER stage to the relation classification phase. Conversely, for joint extraction models like PRGC and TPLinker, concept knowledge further bolsters the models' capacity to capture semantic correlations between entity pairs. This is particularly beneficial for boosting Recall and ameliorating the omission issues.

### 6.3 Hyperparameter Sensitivity Analysis

#### 6.3.1 Effect of Concept Selection Parameters on RE Performance

To explore the effect of the concept selection range, we set two key parameters during concept enhancement: *topk_concepts* and *concept_sim_topk*. The former limits the candidate concepts per token in the concept memory and attention modules to prevent irrelevant concept noise from degrading the contextual representations. The latter controls the retention of highly relevant concepts when constructing relation similarity features, ensuring the classifier focuses on features highly correlated with entity-pair semantics. We conduct parameter sensitivity experiments on the DuIE dataset to validate these settings, and the results are presented in Table 6.

**Table 6:** RE Results with Different topk_concepts / concept_sim_topk Settings on DuIE

| topk_concepts | concept_sim_topk | P | R | F1 |
|---|---|---|---|---|
| 8 | 2 | 60.51 | 48.75 | 54.00 |
| 8 | 4 | 57.51 | 50.97 | 54.05 |
| 8 | 6 | 57.90 | 51.53 | 54.53 |
| 8 | 8 | 55.74 | **53.54** | 54.62 |
| 8 | 10 | 60.01 | 49.83 | 54.45 |
| 8 | 12 | 59.73 | 48.97 | 53.82 |

| | | | | |
|---|---|---|---|---|
| 2 | 6 | 55.60 | 51.89 | 53.68 |
| 4 | 6 | **60.52** | 50.31 | **54.94** |
| 6 | 6 | 59.47 | 50.92 | 54.86 |
| 4 | 8 | 58.49 | 49.78 | 53.78 |

Both parameters significantly impact performance. Expanding *concept_sim_topk* to 8 (given *topk_concepts=8*) maximizes Recall (53.54%) by uncovering latent relations, though larger values introduce noise. Meanwhile, a narrower *topk_concepts=4* (given *concept_sim_topk=6*) achieves the peak F1-score of 54.94% (+1.3% over baseline) by filtering irrelevant concepts. Interestingly, combining their individual optimal values yields poor results due to strong parameter interactions. Overall, *topk_concepts=4* and *concept_sim_topk=6* provide the best balance between semantic enrichment and noise reduction.

### 6.3.2 Effect of Concept Weights on RE Performance

Beyond the standard NER and RE losses, we introduce a concept alignment loss to better align the textual and conceptual semantic representations. To evaluate its impact on model performance, we experiment with different weight coefficients, as detailed in Table 7.

**Table 7:** RE Results with Different align Weights on DuIE

| align | P | R | F1 |
|---|---|---|---|
| 0.0000 | 58.59 | 49.97 | 53.94 |
| 0.0001 | 57.33 | 51.42 | 54.21 |
| 0.0010 | 58.80 | 51.53 | 54.93 |
| 0.0100 | 57.90 | 51.53 | 54.53 |
| 0.1000 | 59.29 | 50.56 | 54.58 |

Experimental results show that introducing the concept alignment loss (*align > 0*) consistently outperforms the baseline model (*align = 0*). At *align* = 0.0010, the model achieves the highest F1-score of 54.93% (a 0.99% increase over the baseline), indicating that the concept alignment constraint can enhance the consistency between entity representations and concept semantics, thereby improving relation extraction performance.

### 6.3.3 Effect of Concept Parameters on NER Performance

To further investigate how various parameter configurations affect NER performance, we conducted extensive experiments centered on three core hyperparameters: the concept alignment loss weight (*align*), the volume of candidate concepts during the concept representation enhancement phase (*topk_concepts*), and the number of highly relevant concepts retained for relation concept similarity computation (*concept_sim_topk*). Evaluated on the DuIE dataset, the detailed results are summarized in Table 8.

**Table 8:** NER Results with Different Parameters on DuIE

| align | topk_concepts | concept_sim_topk | P | R | F1 |
|---|---|---|---|---|---|
| 0.0000 | 8 | 6 | 74.06 | 72.80 | 73.42 |
| 0.0001 | 8 | 6 | 74.02 | 73.09 | 73.55 |

| | | | | | |
|---|---|---|---|---|---|
| 0.0010 | 8 | 6 | 72.62 | 75.01 | 73.80 |
| 0.0100 | 8 | 6 | 71.67 | 76.01 | 73.77 |
| 0.1000 | 8 | 6 | 74.66 | 73.25 | 73.95 |
| 0.01 | 2 | 6 | 71.98 | 74.23 | 73.09 |
| 0.01 | 4 | 6 | 74.01 | 73.42 | 73.71 |
| 0.01 | 4 | 8 | 74.55 | 71.96 | 73.22 |
| 0.01 | 6 | 6 | 74.09 | 73.51 | 73.80 |
| 0.01 | 8 | 2 | 74.36 | 71.75 | 73.03 |
| 0.01 | 8 | 4 | 74.34 | 72.59 | 73.45 |
| 0.01 | 8 | 8 | 72.91 | 75.19 | 74.03 |
| 0.01 | 8 | 10 | 74.18 | 72.59 | 73.37 |
| 0.01 | 8 | 12 | 75.17 | 71.61 | 73.35 |

NER exhibits strong parameter sensitivity to concept injection. Without alignment (*align = 0*), unconstrained concepts degrade performance below the baseline, whereas *align = 0.1* recovers the F1-score to 73.95%. For concept selection, moderate values (*topk_concepts* from 4 to 8) provide slight gains, with *concept_sim_topk=8* achieving the peak F1-score of 74.03% (+0.45%). However, excessive concepts introduce redundancy, impairing boundary detection.

Consequently, concept enhancement does not universally benefit the NER task. Its effectiveness is jointly governed by the concept alignment intensity, the volume of candidate concepts, and the similarity filtering scope. While optimal parameter configurations can boost NER performance, excessively rigid concept constraints or a surplus of conceptual candidates tend to inject noise, ultimately degrading model performance. In contrast to relation extraction, NER exhibits heightened sensitivity to concept injection, necessitating a delicate balance between semantic enrichment and boundary detection stability.

### 6.4 Experiments in Low-Resource Settings

To further investigate the effectiveness and robustness of the FCA-Attention model in low-resource settings, we construct training sets of varying scales from the CMeIE medical dataset, comprising 1,000, 4,000, 7,000, and 10,000 instances, respectively. The validation and test sets are kept fixed to ensure strict comparability of the experimental results across different training sizes. By artificially limiting the volume of training samples, we emulate realistic medical RE scenarios plagued by data scarcity. The corresponding results are summarized in Table 9.

**Table 9:** Results in Low-Resource Settings on CMeIE

| Model | **Data Scale** | NER | | | RE | | |
|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 |
| Baseline | 1000 | 61.79 | 57.10 | 59.35 | 35.38 | 18.88 | 24.62 |
| FCA-Attention | | **62.84** | **61.41** | **62.11** | **38.60** | **22.67** | **28.57** |
| Baseline | 4000 | 70.71 | **71.44** | **71.07** | **37.33** | 30.59 | 33.63 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| FCA-Attention | | **71.84** | 68.96 | 70.37 | 37.00 | **34.73** | **35.83** |
| Baseline | 7000 | 72.47 | **72.61** | **72.54** | **38.61** | 33.63 | 35.95 |
| FCA-Attention | | **73.01** | 71.91 | 72.45 | 38.14 | **39.18** | **38.65** |
| Baseline | 10000 | **74.15** | 70.02 | 72.02 | 38.14 | 35.91 | 36.99 |
| FCA-Attention | | 73.12 | **72.71** | **72.91** | **38.52** | **40.20** | **39.35** |
| Baseline | 12905 | **74.58** | 71.70 | 73.11 | 39.29 | 37.28 | 38.26 |
| FCA-Attention | | 74.04 | **72.60** | **73.31** | **39.75** | **39.63** | **39.69** |

Experimental results reveal that across training scales of 1,000, 4,000, 7,000, and 10,000 instances, our model's F1-score on the NER task fluctuates relative to the Baseline (+2.77%, -0.70%, -0.09%, and +0.89%), yet maintains a generally positive trajectory. In contrast, the RE task exhibits robust F1 improvements of 3.95%, 2.20%, 2.70%, and 2.36%, respectively—all consistently surpassing the 1.43% marginal gain achieved with the full dataset. This highlights that the FCA-Attention architecture yields substantially greater performance dividends in low-resource regimes. This advantage stems from the inherent difficulty of capturing complex medical entity correlations using sparse annotations alone. By injecting prior conceptual information, our proposed formal concept module and concept attention mechanism effectively bridge this gap, mitigating the lack of semantic richness and the instability of relation extraction typical of low-resource environments.

Simultaneously, we conduct experiments on the general-domain DuIE dataset under both low-resource and full-data regimes. Given the massive volume of the original DuIE dataset, our preliminary experiments sampled a subset from the original training corpus for model validation—a configuration functionally equivalent to a low-resource setting. Those initial results confirmed that our method delivers competitive performance even with limited training samples. To further substantiate the model's efficacy at the full data scale, we construct a complete training set utilizing all remaining data after test sample exclusion, while keeping the validation and test sets strictly static. This setup serves to evaluate the applicability of our method on large-scale open-domain datasets and to deeply analyze the impact of training data volume on overall performance. The corresponding results are detailed in Table 10.

**Table 10:** Results across Different Training Scales on DuIE

| Model | Data Scale | NER | | | RE | | |
|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 |
| Baseline | 20000 | 72.65 | 74.54 | 73.58 | 56.41 | 51.14 | 53.64 |
| FCA-Attention | | 71.67 | 76.01 | 73.77 | 57.90 | 51.53 | 54.53 |
| Baseline | 169135 | 74.53 | 76.34 | 75.42 | 58.56 | 57.44 | 57.99 |
| FCA-Attention | | 74.42 | 77.25 | 75.81 | 58.65 | 59.11 | 58.88 |

The results demonstrate that upon scaling the training set to 169,135 instances, our model maintains its superiority over the Baseline, yielding F1-score gains of 0.39% for NER and 0.89% for RE. Notably, as the training volume expands from 20,000 to 169,135, the Baseline's F1-scores climb from 73.58% to 75.42% (NER) and from 53.64% to 57.99% (RE). In parallel, our FCA-Attention

model exhibits corresponding surges from 73.77% to 75.81% (NER) and from 54.53% to 58.88% (RE). These consistent trajectories confirm that scaling up the training corpus substantially bolsters the model's capacity to capture relational patterns within general-domain contexts.

Synthesizing the findings from both low-resource and full-data regimes, the FCA-Attention model consistently yields superior F1-scores across all training scales, underscoring its exceptional stability and generalizability. Specifically, in data-scarce environments, conceptual knowledge acts as a vital semantic supplement; concurrently, in full-data scenarios, the concept enhancement mechanism continues to facilitate the capture of latent semantic associations between entities, further elevating RE performance. Ultimately, these results confirm that our proposed approach is highly effective not only for low-resource relation extraction but also demonstrates broad applicability across large-scale, general-domain datasets.

### 6.5 Ablation Experiments

As a core innovation, this paper introduces a concept memory module and a concept-aware attention mechanism. By leveraging FCA, we construct a concept knowledge base from the training corpus, and dynamically fuse this conceptual knowledge with BERT contextual representations via the concept-aware attention. Concurrently, concept vectors undergo adaptive optimization through a concept alignment constraint during the end-to-end training process, realizing the synergistic learning of structured symbolic knowledge and neural network parameters. During the relation extraction phase, this mechanism integrates entity-pair concept similarity features into the relation representations to characterize latent semantic associations between entities at the conceptual level. This provides crucial knowledge support for relation classification, thereby ultimately boosting the joint extraction performance.

To verify the specific contributions of the concept-aware attention, concept similarity features, concept alignment loss, and concept library parameterization mechanism (dynamic fusion) to the FCA-Attention model, we conduct an ablation study on the CMeIE dataset. The specific variants are defined as follows:

(1) Baseline: The vanilla BERT baseline model without any concept enhancement.

(2) Concept-Aug: A lightweight augmented model incorporating only the concept memory module and concept-aware attention.

(3) Without-Align: The FCA-Attention model excluding the concept alignment loss.

(4) Without-RelFeat: The FCA-Attention model stripped of concept similarity features.

(5) Static-Library: The FCA-Attention model utilizing a static concept library, where the concept vectors in the memory module are frozen and not adaptively updated during training.

The corresponding experimental results are detailed in Table 11.

**Table 11:** Ablation Results of FCA-Attention on CMeIE

| Model | NER | | | RE | | |
|---|---|---|---|---|---|---|
| | P | R | F1 | P | R | F1 |
| Baseline | 74.19±0.34 | 71.09±0.81 | 72.60±0.56 | 37.92±1.20 | 37.71±0.63 | 37.80±0.44 |

| | | | | | | |
|---|---|---|---|---|---|---|
| FCA-Attention | 73.85±0.28 | **72.85±0.46** | **73.34±0.10** | 38.50±2.07 | **41.12±1.52** | **39.71±0.61** |
| Concept-Aug | 74.84±1.35 | 70.89±2.17 | 72.78±0.53 | **38.66±1.34** | 36.75±0.48 | 38.08±0.20 |
| Without-Align | **75.20±1.11** | 71.22±1.21 | 73.15±0.12 | 38.49±1.86 | 40.87±2.06 | 39.58±0.26 |
| Without-RelFeat | 74.22±0.68 | 71.23±0.72 | 73.03±0.31 | 37.81±1.67 | 38.05±0.81 | 37.98±0.38 |
| Static-Library | 75.01±1.49 | 70.83±1.74 | 72.84±0.25 | 38.20±0.54 | 40.34±1.63 | 39.23±0.71 |

The experimental results on the CMeIE dataset demonstrate that the complete FCA-Attention architecture secures the peak F1-scores across both NER and RE tasks. Relative to the unaugmented Baseline, it achieves F1 gains of 0.74% and 1.91%, respectively. Notably, this improvement is predominantly driven by the Recall rate, suggesting that the full model effectively uncovers more latent entities and relations. This effectively alleviates the omission problem stemming from overly conservative predictions, ultimately elevating the model's overall efficacy. Furthermore, the Concept-Aug variant yields F1 increases of 0.18% and 0.28%, proving that introducing a lightweight concept enhancement—relying solely on the concept memory module and concept-aware attention—is sufficient to boost performance. The effectiveness of this module is further corroborated by the experiments in Table 3 and Table 4, which assess the integration of conceptual knowledge into various data-driven models.

Specifically, removing the concept alignment loss (Without-Align) decreases the F1-scores by 0.19% and 0.13% (for NER and RE, respectively). Discarding the concept similarity features (Without-RelFeat) results in notable F1 drops of 0.31% and 1.73%. Furthermore, utilizing a frozen concept library (Static-Library) leads to performance reductions of 0.50% and 0.48%. These ablations reveal distinct mechanism contributions: the concept similarity features serve as the primary driver for RE performance gains, whereas the parameterization of the memory module and the adaptive optimization of concept vectors during training are the main contributors to NER improvements. Finally, the concept alignment loss further elevates overall performance by enforcing semantic consistency between textual and conceptual representations, thereby maximizing the utility of the injected knowledge.

In summary, the ablation study confirms that every module within the FCA-Attention framework plays a distinct and positive role. Specifically, the concept-aware attention seamlessly integrates the concept memory module to enhance token representations; the concept similarity features fortify relation discrimination for entity pairs; the concept alignment loss enforces semantic consistency between textual and conceptual spaces; and the concept memory parameterization mechanism boosts the adaptability of concept knowledge to downstream tasks. The fact that the full model achieves peak performance underscores the complementary nature of these various concept enhancement mechanisms, which work in synergy to jointly elevate the model's representational and discriminative capabilities.

### 6.6 Case Analysis

The joint relation extraction model based on BERT-CRF is a pipeline framework that optimizes the two subtasks of entity recognition and relation extraction simultaneously through joint loss calculation while sharing bottom-level parameters. First, the model uses BERT to obtain the deep

contextual semantic representations of the input text. Then, it performs sequence labeling through the CRF layer to determine entity boundaries and entity types. Next, based on the entity recognition results, it constructs the entity pair representations, concatenates the entity features with the sentence-level contextual features, and inputs them into a fully connected layer for relation prediction.

Based on the BERT-CRF joint extraction framework, the proposed FCA-Attention model introduces a concept enhancement mechanism driven by Formal Concept Analysis. The model constructs a dynamic concept library to structure entities, relations, and features into concept representations. By adding a concept-aware attention layer to the BERT encoder, the text representations can dynamically interact with the concept memory, ensuring that each token contains both textual and conceptual contexts. In the entity recognition task, the model not only relies on the inherent sentence information but also integrates the entity-relation structures from the concept library. In the relation extraction stage, the model incorporates the calculation of entity-pair concept similarities. The combination of entity semantic features and relation concept similarity features further improves model performance. Through the combined action of the concept memory module and the concept-aware attention layer, the FCA-Attention model can retain BERT's context modeling ability while enhancing the capture of entity categories, relation patterns, and semantic structures, thus improving the accuracy and interpretability of the joint extraction task.

To intuitively demonstrate the performance of the FCA-Attention model in medical entity and relation extraction tasks, we select a typical medical text snippet for a granular breakdown analysis, with the results presented in Table 12. This specific instance comprehensively illustrates the complete end-to-end pipeline, tracing the exact transformation from the raw medical text input to the final relation triplet outputs.

**Table 12:** Case Analysis of Medical Text Entity and Relation Extraction using the FCA-Attention Model

| Processing Stage | Case Output |
|---|---|
| Text Input | "The patient has had coronary heart disease for years; chest pain worsened over the past week; taking aspirin." /“患者冠心病多年，近一周胸痛加重，服用阿司匹林” |
| Text Representation | Tokenizes the raw Chinese text and inputs it into BERT to obtain the contextual representation $H$ for each token. |
| Concept Attention Enhancement | Uses token representations as Queries and concept vectors as Keys/Values. Selects the top 8 relevant concepts per token to compute the weight matrix concept_attn, fusing it linearly with $H$ to yield the enhanced representation $H^c$. |
| NER | Decodes $H^c$ into explicit entities:Coronary heart disease (Disease)/冠心病（疾病）,<br>Chest pain (Symptom) / 胸痛（症状）,Aspirin (Drug) / 阿司匹林（药物） |
| Entity Pair Construction | Constructs all directed pairs:冠心病→胸痛,冠心病→阿司匹林, 胸痛→阿司匹林,胸痛→冠心病,阿司匹林→冠心病,阿司匹林→胸痛,etc |
| Base Relation Features | Formulates and concatenates base classification features: head/tail representations, interaction features, difference features, and entity distance features. |
| Concept Similarity Features | Calculates mean concept attention within head/tail entity spans, selects the top 6 concepts for each, computes the entity-pair concept similarity concept_sim, and concatenates it with base features. |
| Relation Prediction | The classifier outputs rel_logits per relation type, scaled to a probability score via sigmoid. Top examples: |

| | |
|---|---|
| RE | [冠心病，药物治疗，阿司匹林]，score=0.6950<br>[冠心病，临床表现，胸痛]，score=0.5270<br>[胸痛，药物治疗，阿司匹林]，score=0.2502<br>[冠心病，相关（导致），胸痛]，score=0.0249<br>[冠心病，预防，阿司匹林]，score=0.0081<br>[冠心病，辅助治疗，阿司匹林]，score=0.0070<br>[冠心病，治疗后症状，胸痛]，score=0.0005……<br>1、Applying schema_mask constraints, such as:<br>“疾病##药物”：药物治疗<br>“疾病##症状”：临床表现、治疗后症状……<br>to filter out invalid entity-relation combinations, yielding:<br>[冠心病，药物治疗，阿司匹林]，score=0.6950<br>[冠心病，临床表现，胸痛]，score=0.5270<br>[冠心病，治疗后症状，胸痛]，score=0.0005<br>applying a probability threshold (i.e., score ≥ 0.5), yielding:<br>[冠心病，药物治疗，阿司匹林]，score=0.6950<br>[冠心病，临床表现，胸痛]，score=0.5270 |
| Final Output | [冠心病（疾病），药物治疗，阿司匹林（药物）]<br>[冠心病（疾病），临床表现，胸痛（症状）] |

### 6.7 Model Interpretability

To further demystify the reasoning mechanism of the proposed FCA-Attention model in the joint NER and RE task, we conduct an interpretability analysis from a multi-grained perspective. Unlike traditional BERT-CRF joint extraction models that rely exclusively on contextual semantic representations, our model introduces a formal concept lattice derived from the training corpus. This lattice abstracts entities and their relation patterns into discrete concept nodes, uniquely characterized by their intensional relations and extensional entities. Subsequently, through the concept-aware attention mechanism and relation concept similarity features, this structured knowledge actively guides the entity recognition and relation classification processes.

#### 6.7.1 Interpretability Analysis of Global Concept Lattice

Initially, we analyze the contextual distribution of entities within relation triplets across the training set to construct relation signatures for both subject and object entities. Specifically, the subject entity signature characterizes the set of relations associated with a head entity, while the object entity signature delineates the relations tied to a tail entity. This design supplies crucial global prior evidence for distinguishing head and tail entities during relation extraction, thereby effectively mitigating the head-tail inversion problem.

During the concept construction phase, entities sharing identical or similar relation signatures are aggregated into formal concepts. Specifically, the intension of a concept denotes a specific set of relation patterns, while its extension encapsulates the collection of entities satisfying these patterns. Essentially, a concept can be interpreted as "a category of entities exhibiting similar relational behaviors." For instance, if multiple disease entities jointly participate in relations such as "clinical

manifestation," "therapeutic drug," and "complication" within the training corpus, they are likely to be subsumed under a unified subject concept. Conversely, if numerous symptom entities collectively serve as descriptive targets for particular diseases or syndromes, they may be clustered into a single object concept. Consequently, the concept lattice not only delineates the similarity-based partitioning of entities but also explicitly uncovers their distinct role functionalities within the relation extraction task.

During the concept pruning phase, we systematically prune redundant concepts by evaluating both their extensional and intensional similarities, selectively preserving those characterized by "richer relations and more focused entities." This operation effectively mitigates the interference of highly similar concepts on the model's attention distribution, rendering the concept library more compact and distinct. Furthermore, our implementation explicitly flags root concepts and overly broad nodes, excluding them by default during the entity-to-concept mapping. This constraint prevents over-generalized concepts from participating in the enhancement mechanism, thereby guaranteeing that the model's attention is exclusively anchored on informative concepts with strong discriminative power.

Ultimately, interpretability at the global concept lattice level is manifested as follows: the FCA-Attention model first induces a global conceptual structure of entity relational behaviors from the training corpus. It then explicates the sources of conceptual knowledge utilized by the model through dimensions such as concept intension, extension, type purity, hierarchical relations, and subject/object role divisions. This overarching structure demonstrates that the model's concept enhancement is anchored in a solid empirical foundation and a rigorous semantic organization. Consequently, it lays a robust groundwork for the subsequent micro-level analyses, including token-level attention interpretation, entity-level recognition interpretation, and relation-level directional judgment interpretation.

### 6.7.2 Interpretability Analysis of Concept Node

A concept node operates as a localized knowledge unit within the global concept lattice. Its extension denotes the set of entities subsumed under the concept, while its intension represents the underlying relation patterns shared by these entities. Consequently, an individual concept node successfully elucidates why the model perceives a particular class of entities as similar, alongside the specific functional roles these entities typically assume in the relation extraction task.

Using node subject_908 as a case study, its details are summarized as follows:


```
{
  "id": "subject_908",
  "kind": "subject",
  "extent":["急性感染性喉炎","肺炎球菌性口炎"],
  "intent":["发病部位","多发季节","药物治疗","辅助治疗"],
  "extent_size": 2,
  "intent_size": 4,
  "dominant_types":["疾病"],
  "type_purity": 1.0,
```

"score": 6.591674

}

The kind attribute of this node is designated as subject, signifying that it is a subject concept; i.e., it primarily characterizes the relational behaviors of head entities. Its extension encompasses two entities: "急性感染性喉炎" (acute infectious laryngitis) and "肺炎球菌性口炎" (pneumococcal stomatitis), both of which are categorized as disease entities in the training corpus. Meanwhile, its intension incorporates four relation types: "发病部位" (site of onset), "多发季节" (peak season), "药物治疗" (drug therapy), and "辅助治疗" (adjuvant therapy). This demonstrates that these two disease entities share highly similar relation patterns when acting as subjects. In essence, the model can interpret this node as a disease-specific subject concept characterized by analogous diagnostic and therapeutic relation patterns.

Ultimately, the interpretability of subject_908 can be encapsulated as follows: this node abstracts "急性感染性喉炎" (acute infectious laryngitis) and "肺炎球菌性口炎" (pneumococcal stomatitis) into a unified disease-centric subject concept characterized by analogous relational behaviors. Their shared relation patterns are explicitly manifested through "发病部位" (site of onset), "多发季节" (peak season), "药物治疗" (drug therapy), and "辅助治疗" (adjuvant therapy). Consequently, if the model heavily attends to this node during inference, it signifies that its entity typing and relation directionality judgments are actively steered by this disease subject concept. This dynamic equips the FCA-Attention framework with a highly traceable conceptual provenance for its prediction pipeline.

### 6.7.3 Interpretability Analysis of Concept-Aware Attention

At the token level, the FCA-Attention model employs concept-aware attention to select the top $k$ most relevant concepts for each token. Specifically, it formulates the token representations as Queries and the concept vectors as Keys/Values, calculating the attention scores between each token and all candidates in the concept library to derive the concept attention distribution, concept_attn. To mitigate the interference of noisy concepts, we retain exclusively the top $k$ concepts with the highest attention weights for conceptual enhancement. To illustrate this mechanism, we examine selected tokens from the sentence "患者冠心病多年，近一周胸痛加重，服用阿司匹林" (The patient has had coronary heart disease for years; chest pain worsened over the past week; taking aspirin) to demonstrate their attention distributions. For brevity, we display only the top two concepts per token, as presented in Table 13.

**Table 13:** Token-Level Top-2 Concept Attention Examples

| Token | Rank | Concept_Id | Attention | dominant_types | type_purity |
|---|---|---|---|---|---|
| 冠 | 1 | subject_1037 | 0.4262 | disease | 1 |
| 心 | 1 | subject_1037 | 0.4559 | disease | 1 |
| 病 | 1 | subject_1037 | 0.3874 | disease | 1 |
| 冠 | 2 | subject_908 | 0.3594 | disease | 1 |
| 心 | 2 | subject_908 | 0.3751 | disease | 1 |
| 病 | 2 | subject_908 | 0.3302 | disease | 1 |

Table 13 illustrates the token-level *top-2* concept attention. Notably, the three tokens of "冠心病"

(coronary heart disease) overwhelmingly attend to disease-based subject concepts. This proves that concept-aware attention not only injects external knowledge into token embeddings but also explicitly visualizes the model's conceptual focus during encoding. Furthermore, when an entity's constituent tokens strongly attend to high-purity concepts, it reveals a clear typing bias already established in the base representation layer, thereby providing a robust foundation for the final NER predictions.

### 6.7.4 Interpretability Analysis of NER

For entity-level interpretability, the model performs NER using concept-enhanced token representations. To assess this knowledge's impact, we analyze the decision process via token-level attention. As shown in Table 13, tokens comprising "冠心病" (coronary heart disease) heavily attend to disease-based subject concepts. This reveals that during boundary delineation and entity typing, the model leverages disease-specific priors from the lattice, naturally biasing its prediction toward a disease entity. Ultimately, concept-aware attention not only boosts the discriminative power of representations but also provides explicit rationales for entity typing and relation directionality, significantly mitigating misclassification errors.

### 6.7.5 Interpretability Analysis of Concept Similarity Features for Relations

For relation extraction, traditional baselines like BERT-CRF build relation features relying entirely on contextual semantics. In contrast, our FCA-Attention model innovatively injects an entity-pair concept similarity feature. By computing the similarity between head and tail entities via their distribution in the concept lattice, this feature explicitly gauges their structural correlation within the underlying conceptual space.

Algorithmically, the model pairs recognized entities and identifies tokens within each head/tail span to evaluate concept_attn. The entity-level concept distribution is then generated by averaging the token-level attention weights for each concept across the entire entity span. For instance (see Table 13), the three tokens of "冠心病" (coronary heart disease) exhibit attention weights of 0.4262, 0.4559, and 0.3874 toward subject_1037, averaging to an entity-level score of 0.4232. Following this logic, subject_908 receives a score of 0.3549, ultimately yielding the full conceptual distribution. To prevent noisy or weakly correlated concepts from misleading the relation classifier, we filter the distribution to retain only the top-k concepts for each head and tail entity, followed by a strict normalization step.

Following this, we compute a concept similarity matrix spanning the top-k concepts of the head and tail entities. This metric is rooted in their intension sets: the similarity between two concepts is measured via the Jaccard similarity coefficient (i.e., the intersection over union of their encompassed relation sets). Naturally, a larger overlap in relations yields a higher similarity score. As a toy example, if Concept 1 contains {"临床表现" (clinical manifestation), "病因" (etiology), "发病部位" (site of onset)} and Concept 2 contains {"临床表现", "发病部位", "药物治疗" (drug therapy)}, the union consists of 4 unique relations and the intersection contains 2. Thus, their Jaccard similarity is 0.5.

Ultimately, the final weighted concept similarity is derived from the normalized head/tail attention weights (head_vals, tail_vals) and the concept similarity matrix (vals_sub). The formulation is given by

$$concept_sim = headvals \times sim_sub \times tailvals \tag{8}$$

The resulting concept_sim feature explicitly quantifies the structural correlation of the entity pair in the conceptual space. It is then concatenated with the head/tail entity representations, alongside the entity interaction, difference, and distance features. This consolidated vector is fed directly into the relation classifier to aid in predicting the underlying relation of the entity pair.

#### 6.7.6 Interpretability Analysis of RE

In the RE phase, all directed entity pairs are constructed from the recognized entities. For each candidate pair, relation features enriched with concept representations for both head and tail entities are fed into a relation classifier. The classifier yields a raw prediction score (rel_logits) for each relation type, which is subsequently transformed into a normalized relation probability (score) via a Sigmoid function.

During relation decoding, a schema_mask is first applied to enforce type constraints. Specifically, the pair "疾病##药物" (Disease##Drug) restricts outputs to relations like "药物治疗" (drug therapy), while "疾病##症状" (Disease##Symptom) strictly limits valid outputs to "临床表现" (clinical manifestation) and "治疗后症状" (post-treatment symptom). Finally, the relation candidates are filtered against a relation_threshold, retaining only those whose score satisfies this minimum confidence criterion.

Using the sentence "患者冠心病多年，近一周胸痛加重，服用阿司匹林" (The patient has had coronary heart disease for years; chest pain worsened over the past week; taking aspirin) as a case study: the model identifies three entities ("冠心病", "胸痛", "阿司匹林") to form 6 directed pairs. The classifier computes raw logits for 264 combinations (6 pairs times 44 relations), converting them to confidence scores via Sigmoid. The schema_mask filters out schema-violating combinations, leaving 3 candidate triplets. Applying the confidence threshold ($score \geqslant 0.5$) yields the final two relational triplets: ("冠心病"[coronary heart disease], "药物治疗"[drug therapy], "阿司匹林"[aspirin]) and ("冠心病"[coronary heart disease], "临床表现"[clinical manifestation], "胸痛"[chest pain]).

### 6.8 Data and Knowledge Fusion Discussion

#### 6.8.1 Effect of Concept Scale

Concept scale acts as a double-edged sword for model performance. While an expansive library broadens entity coverage and supplies rich external knowledge, excessive scale risks injecting low-frequency, redundant, or loosely correlated concepts. This noise causes attention dilution, ultimately impairing the model's discriminative power. Consequently, the optimal concept scale is not strictly maximized but must be rigorously balanced against the training data size and model capacity. To investigate this dynamic, we executed multiple comparative experiments using the low-resource datasets and concept libraries defined previously, with results summarized in Table 13.

**Table 13:** Impact of Concept Scale on Model Performance

| Data Scale | Concept Scale | NER | | | RE | | |
|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 |
| 4000 | 12905 | 71.78 | 69.77 | 70.76 | 36.68 | 34.48 | 35.55 |
| 10000 | 12905 | 71.90 | 74.15 | 73.01 | 37.84 | 39.25 | 38.53 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12905 | 7000 | 73.60 | 71.99 | 72.78 | 41.03 | 39.29 | 40.14 |
| 12905 | 4000 | 74.05 | 72.59 | 73.31 | 40.62 | 39.71 | 40.16 |

Results indicate that for training sets of 4,000 and 10,000 samples, expanding the concept repository to 12,905 degrades precision, recall, and F1 compared to identically scaled libraries (with F1 dropping by 0.28% and 0.82%, respectively). In contrast, given a training set of 12,905 samples, restricting the concept library to 7,000 or 4,000 robustly improves the F1 score by roughly 0.46%.

We attribute this to diminishing returns: while new concepts expand entity coverage, they simultaneously inject low-quality, over-generalized, and unstable representations. This causes the enhancement process to devolve from "semantic supplementation" into "noise interference." Conversely, the performance gains achieved by shrinking the library explicitly confirm the abundance of noise in the raw data, highlighting the absolute necessity of rigorous concept filtering to build a high-quality, low-noise repository.

### 6.8.2 Effect of Concept Quality

To extract high-quality core concepts, we introduce three criteria: Extent Size (entity count), Intent Size (relational feature count), and Extent Type Purity (entity type consistency). The optimal concepts exhibit a moderate extent size, well-defined relational intents, and high type purity. Rationally, excessively small extents capture spurious noise rather than stable patterns, while overly large extents suffer from over-generalization and diminished discriminative power. Meanwhile, a larger intent imposes stricter relational constraints, and high purity prevents noise from heterogeneous entity mixing.

However, since FCA-Attention's enhancement mechanism depends on entity-to-concept mappings, filtering refines the repository but down-scales its volume and reduces entity coverage, risking performance degradation. Ultimately, optimizing this filtering process demands a strict trade-off between "concept quality" and "concept scale."

To evaluate how concept quality dictates model performance, we build an initial concept library from the DuIE and CMeIE training sets and apply a rigorous core concept filtration process. Table 14 outlines the resulting experimental evaluation.

**Table 14:** Performance Evaluation of Concept Filtration

| Dataset | Data Scale | Filtering | NER | | | RE | | |
|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 |
| DuIE | 20000 | without | 71.67 | 76.01 | 73.77 | 57.90 | 51.53 | 54.53 |
| | | with | 72.92 | 75.76 | 74.32 | 59.62 | 51.78 | 55.43 |
| CMeIE | 12905 | without | 74.04 | 72.60 | 73.31 | 39.75 | 39.63 | 39.69 |
| | | with | 75.90 | 69.56 | 72.59 | 41.67 | 38.18 | 39.85 |

As revealed by the results, core concept filtering outclasses the unfiltered raw repository across both benchmarks. On DuIE, the F1 score advances from 54.53% to 55.43% (up 0.90%), and on CMeIE, it grows from 39.69% to 39.85% (up 0.16%). This performance leap proves that not all concepts contribute positively to relation extraction. Ultimately, filtering out low-quality concepts effectively shields the model's representation learning and relation classification from noisy concept interference.

### 6.8.3 Effect of Concept Scale under High-Quality Concept Constraints

To evaluate how expanding a high-quality concept repository dictates performance in low-resource settings, we hold the training set size constant. Specifically, the previously defined 4,000 CMeIE samples form the training set to build the initial library. We then filter an additional 1,000 and 2,000 samples from the original 12,905 pool solely to scale up the concept library, explicitly withholding them from supervised training. This isolates the impact of concept repository size on the FCA-Attention model. Table 15 summarizes the experimental results.

**Table 15:** Performance Impact of High-Quality Concept Scale Expansion

| Data Scale | Concept Scale | Filtering | NER | | | RE | | |
|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 |
| 4000 | 0 | / | 70.71 | 71.44 | 71.07 | 37.33 | 30.59 | 33.63 |
| 4000 | 4000 | without | 71.84 | 68.96 | 70.37 | 37.00 | 34.73 | 35.83 |
| 4000 | 4000 | with | 71.10 | 70.33 | 70.71 | 36.73 | 34.12 | 35.38 |
| 4000 | 5000 | with | 72.07 | 70.40 | 71.22 | 40.12 | 33.19 | 36.33 |
| 4000 | 6000 | with | 72.49 | 70.33 | 71.39 | 38.08 | 33.58 | 35.69 |

Further comparison reveals that at a concept library size of 4,000, filtering yields an F1 score of 35.38%—outperforming the no-enhancement baseline by 1.75% but marginally trailing the unfiltered baseline (35.83%). This implies a clear trade-off: while filtering purifies the repository, it can inadvertently purge concepts vital for entity coverage, causing a coverage drop that penalizes model performance.

Expanding the concept repository from 4,000 to 5,000 yields optimal performance, pushing precision, recall, and F1 to 40.12%, 33.19%, and 36.33%—outperforming the no-enhancement baseline by 2.70% and the 4,000-scale filtered baseline by 0.95%. This underscores that a well-calibrated expansion of high-quality concepts broadens semantic coverage, equipping more entities with valuable contextual priors to drive performance gains.

However, further expanding the repository to 6,000 triggers a performance drop, with the F1 score sliding to 35.69% (down from 36.33% at 5,000). This underscores that scaling is bounded by diminishing returns. While further expansion broadens entity coverage, it inevitably injects redundant, loosely coupled, or out-of-distribution concepts. This noise dilutes the concept attention weights, ultimately eroding the classifier's discriminative power during relation identification. Hence, optimizing the repository requires a strict trade-off between concept coverage and concept quality.

In summary, the integration of a concept repository significantly boosts model performance, while a well-calibrated expansion of high-quality concepts can further augment these gains. However, when the repository scale becomes excessive, newly incorporated concepts risk injecting noise and redundant information, thereby bottlenecking or even degrading performance. Consequently, the efficacy of the proposed FCA-Attention model is not merely a function of repository size; rather, it is intricately intertwined with concept quality, entity coverage, and the semantic alignment between newly introduced concepts and the target task.

## 7.Conclusions and Future Work

### 7.1 Research Conclusions

This paper proposes FCA-Attention, a relation-aware attention model that incorporates FCA for the joint extraction of entities and relations from medical text. By constructing a structured relational formal concept repository and designing a novel concept-aware attention mechanism, the model deeply integrates symbolic conceptual knowledge into the BERT encoding process, thereby achieving a precise semantic alignment between textual representations and the conceptual space.

Experimental results on the CBLUE medical benchmark dataset demonstrate that FCA-Attention significantly outperforms existing baseline models in both entity recognition and relation extraction tasks, with performance gains being particularly pronounced in low-resource scenarios. Furthermore, evaluations on the DuIE general-domain dataset corroborate the model's effectiveness in general-domain entity recognition and relation extraction, manifesting robust cross-domain generalization capability.

Ablation experiments validate the effectiveness of each constituent module, while case studies vividly demonstrate the model's inherent interpretability. Furthermore, the discussion results confirm that model performance is intricately tied to both concept quality and repository scale. Rationally constructing a high-quality concept repository and enhancing entity coverage are instrumental in fully leveraging the facilitating role of conceptual knowledge in entity representation and relational semantic modeling, ultimately boosting overall model performance.

### 7.2 Research Limitations

Although the FCA-Attention model proposed in this study has achieved superior performance over baseline models in both Chinese medical named entity recognition and relation extraction tasks, and has exhibited robust generalization capabilities across the general data domain, certain limitations still remain to be acknowledged.

Firstly, model performance is intricately intertwined with both concept quality and repository scale. While high-quality concepts provide precise and effective semantic knowledge, a repository contaminated with excessive low-quality or redundant concepts inevitably introduces noise, thereby attenuating the efficacy of concept enhancement; thus, rigorous filtering and optimization during the concept construction phase are a prerequisite. However, because the FCA-Attention model inherently relies on entity-to-concept mapping pathways to implement enhancement, the repository contraction during filtration prevents certain entities from matching any corresponding concepts. This erodes the overall entity coverage and subsequently undermines model performance. Consequently, a delicate trade-off exists between maximizing concept quality and maintaining robust concept coverage capability.

Secondly, the FCA-Attention model utilizes its concept-aware attention mechanism to select the top-k most relevant concepts for token-level representation enhancement. During the relation extraction phase, it retrieves the top-k concepts for both the head and tail entities, computing entity-level concept similarity features to assist relation classification. However, if the parameter k is configured too small,

the model struggles to thoroughly mine the latent semantic knowledge of entities, yielding fragmented or incomplete concept representations. Conversely, an excessively large k introduces weakly correlated or completely irrelevant concepts, intensifying noise interference, diluting the discriminative power of concept representations, and ultimately compromising the efficacy of relation features. Consequently, a critical trade-off persists between maximizing concept information utilization and maintaining rigid noise control.

In conclusion, while this study has empirically verified the correlations between model performance and concept quality, repository scale, and the concept quantity parameter k, several optimization pathways remain open. Specifically, how to dynamically weigh the relationships between concept quality, entity coverage, and repository scale, as well as the interplay between full utilization of conceptual information and rigid noise control, requires further exploration. Achieving the global optimal balance among concept utilization efficiency, knowledge quality, and model performance remains an important avenue for subsequent work.

### 7.3 Future Work

Future research will primarily focus on the following dimensions: First, we aim to explore the dynamic updating mechanisms for the concept repository, enabling conceptual knowledge to evolve and expand synchronously with evolving clinical practices. Secondly, we plan to extend the current framework to encompass multimodal medical data. Thirdly, we will investigate the application of concept-aware attention in knowledge distillation and lightweight model deployment. Finally, comprehensive evaluations will be conducted through large-scale validation within actual clinical systems to verify the practical efficacy of the proposed model.

## Data availability statement

The datasets used in this study, including CMeIE from the CBLUE benchmark and DuIE, are publicly available from their respective official repositories. No new clinical data were collected or generated in this study.

## Ethics statement

This study did not involve the direct recruitment of human participants or the collection of new clinical data. All experiments were performed using publicly released benchmark datasets, including CBLUE/CMeIE and DuIE, in accordance with their stated access and usage conditions. This study did not involve the direct recruitment of human participants or the collection of new clinical data. All experiments were performed using publicly released benchmark datasets, including CBLUE/CMeIE and DuIE, in accordance with their stated access and usage conditions.

## Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## CRediT authorship contribution statement

**Chunlei Cheng:** Conceptualization, Formal analysis, Methodology, Funding acquisition, Project administration, Writing-review & editing. **Shujun Tan:** Formal analysis, Validation, Writing-original draft, Investigation. **Xinyi Xiong:** Investigation, Resources. **Cong Cheng:** Investigation, Resources.

## Funding Statement

This work was supported by the Natural Science Foundation of Jiangxi Province (Grant No. 20224BAB206102) and the National Natural Science Foundation of China (Grant No. 82260988).

## References


[1] Wang, Ling,Hao, Haoyu,Yan, Xue, et al.From biomedical knowledge graph construction to semantic querying: a comprehensive approach[J].SCIENTIFIC REPORTS,2025,15:1-13.

[2] Priyankar B ,Sriram S ,C. W S , et al.A Survey on Recent Named Entity Recognition and Relationship Extraction Techniques on Clinical Texts[J].Applied Sciences,2021,11(18):8319-8319.

[3] Cai, Linkun,Li, Jia,Lv, Han, et al.Integrating domain knowledge for biomedical text analysis into deep learning: A survey[J].Journal of Biomedical Informatics,2023,143.

[4] Lu Zhou,Chao-Yong Wu,Xi-Ting Wang, et al.Traditional Chinese Medicine Synonymous Term Conversion:A Bidirectional Encoder Representations from Transformers-Based Model for Converting Synonymous Terms in Traditional Chinese Medicine[J].World Journal of Traditional Chinese Medicine,2023,9(2):224-233.

[5] Khetan, Vivek,Rizvi, Md Imbesat Hassan,Huber, Jessica, et al.MIMICause: Representation and automatic extraction of causal relation types from clinical notes[C].FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (ACL 2022).Laos,2022:764-773.

[6] Lu, Yonghe,Chen, Hongyu,Zhang, Yueyun, et al.Research on entity relation extraction for Chinese medical text[J].Health Informatics Journal,2024,30(3).

[7] Kong, Wenjun,Xia, Yamei.CARE: Co-Attention Network for Joint Entity and Relation Extraction[C].2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC-COLING 2024 - Main Conference Proceedings.Norway,2024:2864-2870.

[8] Su, Yansen,Wang, Minglu,Wang, Pengpeng, et al.Deep learning joint models for extracting entities and relations in biomedical: a survey and comparison[J].Briefings in Bioinformatics,2022,23(6).

[9] Xu, Zekai,Tong, Jijun,Tong, Yetao, et al.URRel: Joint Entity and Relation Extraction in Chinese Medical Texts via Unified Representation and Reinforcement Learning With Overlapping Relations[J].Expert Systems,2026,43(3).

[10] Min, Bonan,Ross, Hayley,Sulem, Elior, et al.Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey[J].ACM Computing Surveys,2024,56(2):1-40.

[11] Hsu, Enshuo,Roberts, Kirk.Leveraging large language models for knowledge-free weak supervision in clinical natural language processing[J].SCIENTIFIC REPORTS,2025,15:1-10.

[12] Khatri, Minal,Yin, Yanbin,Deogun, Jitender.Enhancing Interpretability in Medical Image Classification by Integrating Formal Concept Analysis with Convolutional Neural Networks[J].Biomimetics (Basel, Switzerland),2024,9(7):421-421.

[13] Tiaiba, Hafida,Sabri, Lyazid,Kazar, Okba.Advanced prediction of events and temporal expressions in medical text using the Jena API: integrating ontologies and deep learning[J].Turkish Journal of Electrical Engineering and Computer Sciences,2025,33(1):65-85.

[14] Wang, Weijie,Li, Xiaoying,Ren, Huiling, et al.Chinese Clinical Named Entity Recognition From Electronic Medical Records Based on Multisemantic Features by Using Robustly Optimized Bidirectional Encoder Representation From Transformers Pretraining Approach Whole Word Masking and Convolutional Neural Networks: Model Development and Validation[J].JMIR MEDICAL INFORMATICS,2023,11:e44597.

[15] Mienye, Ibomoiye Domor,Swart, Theo G.,Obaido, George.Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications[J].INFORMATION,2024,15(9):517.

[16] Lee, Jinhyuk,Yoon, Wonjin,Kim, Sungdong, et al.BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J].Bioinformatics,2020,36(4):1234-1240.

[17] Jin, Myeong,Choi, Sang-Min,Kim, Gun-Woo.COMCARE: A Collaborative Ensemble Framework for Context-Aware Medical Named Entity Recognition and Relation Extraction[J].ELECTRONICS,2025,14(2).

[18] Kosprdic, Milos,Prodanovic, Nikola,Ljajic, Adela, et al.From zero to hero: Harnessing transformers for biomedical named entity recognition in zero- and few-shot contexts[J].Artificial Intelligence in Medicine,2024,156.

[19] Wang, Yucheng,Yu, Bowen,Zhang, Yueyang, et al.TPLinker: Single-stage Joint Extraction of Entities and Relations Through Token Pair Linking[C].COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference.Niger,2020:1572-1582.

[20] Zheng, Hengyi,Wen, Rui,Chen, Xi, et al.PRGC: Potential relation and global correspondence based joint relational triple extraction[C].ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference.Niger,2021:6225-6235

[21] Yang, Zhe,Huang, Yi,Feng, Junlan.Learning to Leverage High-Order Medical Knowledge Graph for Joint Entity and Relation Extraction[C].FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (ACL 2023).Andorra,2023:9023-9035.

[22] Gligorijević F M ,Bogdanović M ,Puflović D , et al.Long-Term Categorization Capability Assessment of Semantically Reduced Concept Lattices[J].Applied Sciences,2026,16(11):5490-5490.

[23] J. Xia, C. Cheng, H. Chen, et al., "Chinese entity relation extraction strategy based on predicate concept connectivity," Journal of Frontiers of Computer Science and Technology, vol. 8, no. 11, pp. 1345-1357, 2014. (in Chinese)

[24] C. Cheng, J. Zou, Q. Ye, et al., "End-to-end triplet extraction integrating relational formal concepts," Computer Engineering and Applications, vol. 59, no. 9, pp. 182-189, 2023. (in Chinese)

[25] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[26] Gu, Yu,Tinn, Robert,Cheng, Hao, et al.Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing[J].ACM Transactions on Computing for Healthcare,2022,3(1):1-23.

[27] Cai, Linkun,Li, Jia,Lv, Han, et al.Integrating domain knowledge for biomedical text analysis into deep learning: A survey[J].Journal of Biomedical Informatics,2023,143.

[28] Liu W, Zhou P, Zhao Z, et al. K-bert: Enabling language representation with knowledge graph[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(03): 2901-2908.

[29] Peters M E, Neumann M, Logan R, et al. Knowledge enhanced contextual word representations[C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). 2019: 43-54.

[30] Wang, Ruize,Tang, Duyu,Duan, Nan, et al.K-ADAPTER: Infusing Knowledge into Pre-Trained Models with

Adapters[C].FINDINGS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, ACL-IJCNLP 2021.The Federal Democratic Republic of Ethiopia,2021:1405-1418.

[31] Zhang, Zhengyan,Han, Xu,Liu, Zhiyuan, et al.ErniE: Enhanced language representation with informative entities[C].ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference.New Caledonia,2020:1441-1451.

[32] Lewis, Patrick,Perez, Ethan,Piktus, Aleksandra, et al.Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks[C].ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 33, NEURIPS 2020.The Federal Democratic Republic of Ethiopia,2020.

[33] Guu, Kelvin,Lee, Kenton,Tung, Zora, et al.REALM: Retrieval-Augmented Language Model Pre-Training[C].INTERNATIONAL CONFERENCE ON MACHINE LEARNING, VOL 119.The Federal Democratic Republic of Ethiopia,2020:3887-3896.

[34] Borgeaud, Sebastian,Mensch, Arthur,Hoffmann, Jordan, et al.Improving Language Models by Retrieving from Trillions of Tokens[C].INTERNATIONAL CONFERENCE ON MACHINE LEARNING, VOL 162.Republic of Moldova,2022:2206-2240.

[35] Zhai S, Qi G, Wang Y, et al. Knowledge Fusion via Bidirectional Information Aggregation[J]. arXiv preprint arXiv:2507.08704, 2025

[36] Guan, Tongfeng,Zan, Hongying,Zhou, Xiabing, et al.CMeIE: Construction and Evaluation of Chinese Medical Information Extraction Dataset[C].Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).Nigeria,2020:270-282.

[37] Li, Shuangjie,He, Wei,Shi, Yabing, et al.DuIE: A Large-Scale Chinese Dataset for Information Extraction[C].Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).Nigeria,2019:791-800.