


## Article

# Knowledge Base Question Answering via Semantic Analysis

Yibo Liu , Haisu Zhang, Teng Zong, Jianping Wu and Wei Dai

School of Information Communication, National University of Defense Technology, Wuhan 430014, China

\* Correspondence: liuyibo@nudt.edu.cn

**Abstract:** Knowledge Question Answering is one of the important research directions in the field of robot intelligence. It is mainly based on background knowledge to analyze users' questions and generate answers. It is one of the important application methods of knowledge graph technology. Compared with the traditional expert system of question and answer, it has the advantage of a large-scale background knowledge base and the traceability and interpretability of the question-answering process. Compared with the current ChatGPT (Chat Generative Pre-trained Transformer) technology, it has advantages in the proprietary segmentation field. Aiming at the problem of the accuracy of existing knowledge question-answering methods being low, this paper studies the method of semantic analysis for knowledge question-answering under the support of a knowledge database, proposes a knowledge question-answering method based on the superposition of multiple neural network models, and conducts experimental verification on the publicly available NLPCC2016KBQA (Knowledge Q&A Tasks in the 2016 Natural Language Processing and Chinese Computing Conference) data set. The experimental results show that the F1 value of this method is higher than that of the baseline model.

**Keywords:** knowledge graph; knowledge question answering; semantic analysis; neural network model



**Citation:** Liu, Y.; Zhang, H.; Zong, T.; Wu, J.; Dai, W. Knowledge Base Question Answering via Semantic Analysis. *Electronics* **2023**, *12*, 4224. <https://doi.org/10.3390/electronics12204224>

Received: 21 September 2023

Revised: 7 October 2023

Accepted: 10 October 2023

Published: 12 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Whether a machine can correctly answer the questions described by human language instead of traditional information service methods has always been the focus and difficulty of research in natural language processing, artificial intelligence, human–computer interaction, and other fields [1]. With the breakthrough of AI in the direction of cognitive intelligence, AI has been able to carry out mature industrial landing in the fields of face recognition, speech recognition, and so on [2]. In contrast, the development of AI in the direction of cognitive intelligence is still relatively slow [3]. Researchers in knowledge Q&A want machines to have certain cognitive, experiential, and thinking abilities like humans. However, there is still a certain gap in achieving this goal at present [4,5]. Cognitive intelligence can be used in fields such as knowledge Q&A or intelligent translation [6]. In recent years, the rapid development of big data technology and knowledge graph technology has led to more and more data being collected, stored [7]. Since the end of 2022, ChatGPT and its competitors have gradually made intelligent Q&A part of people's daily lives. However, ChatGPT's pre-training model parameterizes knowledge and has poor interpretability [8]. In case of inaccurate answers, it is impossible to carry out technical backtracking. Intelligent question-answering technology based on a knowledge graph can use the knowledge base to provide background knowledge and experience for machines, so as to achieve the traceability of question answering [9,10]. This is significant for some proprietary fields with higher requirements. Given this, based on the current research of intelligent question-answering technology based on knowledge graphs, this paper combines deep learning models to further explore intelligent question-answering based on knowledge graphs.

This paper proposes an intelligent question-answering method based on multi-neural-network cooperation for the Chinese language corpus. By identifying the entity of the question and distinguishing the type of question, it generates the query statement of the

knowledge base and, finally, queries the knowledge base to obtain the resulting answer, which can achieve good results on the open NLPCC2016KBQA data set. The contributions of this paper mainly include the following aspects:

- (1) This paper summarizes the current mainstream methods of knowledge Q&A, and compares and describes the characteristics and advantages of each of the four mainstream methods: the rule-based question-answering method, information retrieval question-answering method, ChatGPT method, and semantic analysis question-answering method.
- (2) The published NLPCC2016KBQA data set was optimized. For the public NLPCC2016KBQA, segmentation and data processing were carried out to generate the special data set for question entity recognition and the special data set for question type identification; support model training; extracting the triad in the NLPCC2016KBQA data set; and generating the knowledge graph, as the background knowledge for intelligent question answering.
- (3) A method of multi-neural-network cooperation, the BBCB (BERT-BiLSTM-CRF-BERT) model, is proposed to research the machine intelligence question-answering task. For the entity extraction task, a model was constructed based on BERT-BiLSTM-CRF and experiments were conducted on the constructed question entity recognition data set. The final F1 value was 95.5%. For the question-type judgment task, a BERT-FC model was constructed. On the NLPCC2016KBQA data set, the final F1 value was 99.1%.

The rest of this paper is organized as follows: Section 2 introduces the related work of intelligent question answering. Section 3 introduces the method model proposed in this paper. Section 4 describes the experimental verification of the method, including the experimental environment, data set, and result analysis. Section 5 presents an error analysis of the experimental results. Section 6 summarizes and analyzes the work of this paper.

## 2. Literature Review

The earliest development of machine automatic question answering can be traced back to 1950. Alan Turing [11] proposed that the key to verifying whether a machine is intelligent is to see whether it can correctly answer questions. In 1963, Green [12] first designed the Baseball software to answer questions about the baseball field. This is the earliest automatic question-answering system. In 1966, the Massachusetts Institute of Technology of the United States realized the ELIZA intelligent chat robot [13]. Since 1999, the TREC (Text Retrieval Conference) has been organizing evaluation tasks related to intelligent question answering and has held the longest-running evaluation task of text retrieval.

### 2.1. Mainstream Technology

In terms of technical implementation, there are currently four main types of methods [14,15]: the first is a rule-based question-and-answer method, the second is an information-retrieval-based question-and-answer method, the third is a semantic analysis question-and-answer method, the fourth is a Large Language Model (LLM)-based question-and-answer method, and deep learning technology is widely used in all of these methods [16].

The rule-based approach is the earliest one [17]. Woods et al. [18] used the rule-based approach to build a question-answering system called LUNAR to match users' questions through rules. Liu et al. [19] proposed the use of natural language processing tools to classify questions and the use of rules to match answers. Riloff E. et al. [20] used the rule-matching method to build the Arabic question-answering system. The rule-based method is simple to implement and does not need background knowledge. However, this method can only formulate good rules for limited vertical fields, which cannot be used in non-limited fields, such as open chat robots, and requires a lot of manpower and time to make rule templates.

The method based on information retrieval applies the traditional idea of information extraction to solve the KGQA (Knowledge Graph Question Answer) problem [21]. First, all entity nodes associated with the subject entity of the problem are searched from the

knowledge base as candidate answers; then, the candidate answers are sorted and the optimal solution is selected. Bordes [22] designed a coding model for candidate answers, which generates vectors with equal dimensions from candidate answers and questions, and sorts the candidate answers by the product value. Zhang et al. [23] used an attention mechanism to train and sort different candidate answers. The method based on information retrieval applies to some end-to-end scenarios, and the black-box-style analytical model in the middle makes the whole process less interpretable.

The question-answering method based on semantic analysis often uses machine learning or deep learning models to conduct semantic analysis on question statements and generate query statements for the knowledge base. Ding Jiajie et al. [24] proposed a data enhancement model that can filter the data with low quality. Peng Yu et al. [25] proposed a three-stage question-and-answer model based on the BERT model. Li J et al. [26] built a knowledge graph for the tourism field on this basis. They completed the knowledge Q&A in the tourism field using the BERT model. Shamsabadi A. S. et al. [27] proposed the use of multiple classifiers in the pipeline in the solution for the identification of hidden relationships, where each classifier uses a separate BiGRU neural network. The question-answering method based on semantic analysis can independently learn and generate models with complex parameters and functions. Because it parses natural language into the logic of knowledge base query, the reasoning process is more interpretable and the overall accuracy of question and answer is higher.

The most mainstream product of the LLM (Large Language Model)-based Q&A method is the recent ChatGPT [28]. The performance of ChatGPT in various general Q&A tasks has reached an unprecedented level. However, compared with the knowledge Q&A based on the knowledge atlas, it has disadvantages such as non-interpretability of Q&A results and poor reasoning ability in proprietary fields. In addition, the current ChatGPT code is not open-source. Based on the above analysis, this paper mainly adopts a semantic-analysis-based question-and-answer method to complete knowledge question and answer. We use the dictionary method and multi-neural-network collaboration to analyze topic entities and question types, and search for the final answer based on a knowledge graph. In this process, this paper proposes a question-and-answer method based on the BBCB (BERT-BiLSTM-CRF-BERT) model and background knowledge.

In addition, in the process of knowledge Q&A, entity disambiguation is a difficult problem. Currently, there are various methods to solve entity disambiguation problems, including the well-known first method, multi-feature-based method, etc. Among them, the multi-feature-based disambiguation method has a better effect [29].

## 2.2. Related Data Sets

The realization of machine question answering cannot be separated from the support of an artificial rule base or background knowledge, while the artificial rule method uses experts to formulate rules in specific fields to classify and judge questions, which consumes a lot of manpower and is generally limited to question answering in specific fields [30]. Only a large enough background knowledge can realize intelligent question answering in open fields. In 2012, Google put forward the concept of a knowledge graph [31] and built a large-scale triplet knowledge base using knowledge-engineering-related technologies. After that, knowledge graph technology developed rapidly. At present, many large-scale knowledge graphs have been built [32], including the Google Knowledge graph, DBpedia Knowledge graph, OpenIE Knowledge graph, Wikidata Knowledge graph, Zhishi.me of Southeast University in China, XLORE of Tsinghua University, CN-pedia of Fudan University, etc. The rapid development of knowledge graphs has laid the foundation for intelligent question-answering research in the open field. In this paper, we mainly extract the triad from the Q&A evaluation data set and construct the knowledge graph to provide support for the experiment of the Q&A method. In terms of evaluation data sets, the most famous foreign question-and-answer evaluation data sets include the TREC (Text Retrieval Conference) data set, CLEF (Cross Language Evaluation Forum) data set, and NTCIR (NACSIS Test

Collections for IR) data set. In the Chinese field, in 2016, the NLPCC (The Conference on Natural Language Processing and Chinese Computing) released a large-scale Chinese question-answering data set NLPCC ICCPOL 2016 KBQA [33]; we used this dataset to carry out experimental verification on the two key modules of question entity recognition and question type judgment in our method.

### 3. Method

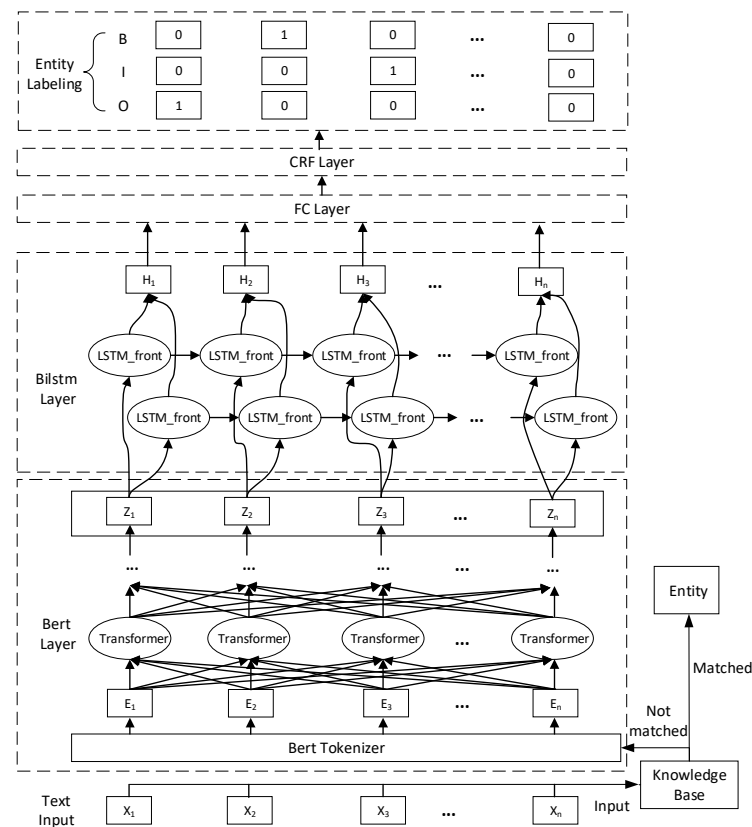
This paper proposes a knowledge question-answering model based on semantic analysis, which is a multi-neural-network collaboration model. The method model in this paper is described from three aspects: question entity recognition, question type determination, and answer generation.

For knowledge Q&A tasks, this paper mainly determines model selection through the following methods: Firstly, due to the strong feature extraction ability of the current BERT model, it was selected as a feature encoder for question entity recognition and question type judgment. Secondly, we chose mature models based on previous research results. Thirdly, the final model combination was determined through experimental verification, which is the optimal model in the experiment.

This method can achieve good results in knowledge Q&A. In terms of contribution, the proposed model can be used for both question entity recognition and question type judgment. This method can lay the foundation for further research in the future.

#### 3.1. An Entity Recognition Model for Question Sentences

In terms of question entity recognition, this model proposes a combination of knowledge base matching and the BERT-BiLSTM-CRF model. The network structure is shown in Figure 1.



**Figure 1.** Question entity recognition based on the background knowledge base and BERT-BiLSTM-CRF model.

With the relatively high stability of high-quality background knowledge, knowledge base matching is first performed for the question text. If the matching is successful, the entity is directly extracted, which can greatly improve the entity recognition efficiency of the model. If the matching fails, input the question text into the BERT–BiLSTM–CRF model and label the entity, obtain the BIO label of the question, and identify the question entity. In the BIO notation, each element is labeled as B, I, or O, representing the beginning, middle, and others of the entity.

### 3.1.1. Question Entity Recognition BiLSTM Layer

LSTM network is a special form of a cyclic neural network, which can better deal with the sequence data similar to the text line [34]. Compared with the traditional cyclic neural network, it can effectively solve the problem of feature resolution when the question text is long. After studying relevant models, this paper uses the BiLSTM model for further feature processing of the output vectors of the BERT model. An example of the node structure of BiLSTM is shown in Figure 2 [35], and the update gate and forget gate in the network structure can effectively control the retention and resolution of information.

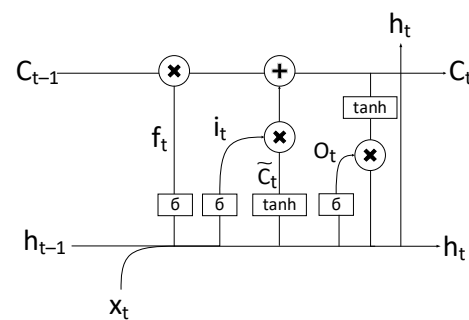


Figure 2. An example of the node structure of BiLSTM.

The forgetting gate is calculated by the following formula:

$$f_t = \sigma(M_f \times [h_{t-1}, x_t] + k_f) \quad (1)$$

The updated door is calculated by the following formula:

$$i_t = \sigma(M_i \times [h_{t-1}, x_t] + k_i) \quad (2)$$

$$\tilde{C}_t = \tanh(M_c \times [h_{t-1}, x_t] + k_c) \quad (3)$$

The output status is calculated by the following formula:

$$O_t = \sigma(W_O[h_{t-1}, x_t] + k_o) \quad (4)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (5)$$

$$h_t = O_t \times \tanh(C_t) \quad (6)$$

In the above formula,  $M_f$ ,  $M_i$ ,  $M_c$ ,  $M_o$ ,  $k_f$ ,  $k_i$ , and  $k_c$  are the weight parameter matrices and  $\tanh$  represents a hyperbolic tangent function.  $X_t$  represents the data to be input,  $h_{t-1}$  represents the input data of  $t - 1$  at the previous time,  $h_t$  represents the input data at the current time, and  $\sigma$  represents a sigmoid function to ensure that the output result is a number between 0 and 1.

### 3.1.2. CRF Layer of Question Entity Recognition

Before feature encoding enters the CRF layer, the FC fully connected layer is used to process the input vector as the dimension required by the CRF layer, and the final

output of the CRF model is the BIO label of the text.  $y$  is the three-dimensional matrix  $[a, b, c]$ , where  $a$  represents the number of data lines in a batch,  $b$  represents the characters corresponding to the line of text, and  $c = 3$  represents the probability of each character corresponding to three types of BIO labeling. The specific formula for the FC layer is shown in Formula (7):  $k$  represents the paranoia matrix,  $i$  represents the input matrix, and  $W$  represents the weight matrix:

$$y = iW^T + k \quad (7)$$

At present, the CRF model is widely used in various entity recognition tasks, including entity recognition of questions. CRF belongs to the discriminant probability undirected graph model. When the upper model is given the input random variables, the CRF model can obtain the conditional probability distribution of the random variables for conditional decoding. Although the BiLSTM model can capture long-distance text features, it cannot sense the dependency between adjacent labels, and CRF can better perceive the relationship between entities and adjacent characters. Therefore, in this paper, after the BERT-BiLSTM model, the CRF model is added to decode the final entity label code. The specific calculation formula is as follows:

$$Score(X, Y) = \sum_{i=1}^n S_{i,y_i} + \sum_{i=0}^n A_{y_i,y_{i+1}} \quad (8)$$

where  $X = (x_1, x_2, \dots, x_n)$  represents the set of input feature vectors,  $Y = (y_1, y_2, \dots, y_n)$  represents the prediction sequence,  $X_i$  represents the feature vector of the  $i$ th character,  $S$  represents the score matrix of the output result of the model before CRF, and  $S_{ij}$  represents the score of the  $j$ th label of the  $i$ th character.

In this paper, the BIO labeling method is used; so, there are three labels in total.  $j$  is taken from the set  $\{1, 2, 3\}$ ,  $A$  refers to the transfer matrix, and  $A_{ij}$  indicates the probability that tag  $i$  will transfer to tag  $j$ . Finally, the higher the score value, the higher the probability of labeling. The CRF model uses the maximum likelihood function to obtain the output sequence of the maximum score as the labeling result.

### 3.2. Attribute Similarity Judgment

Based on question entity recognition, this paper uses the BERT model to match the background knowledge for attribute similarity judgment. The model is shown in Figure 3. For the input question, the attributes in the knowledge base are cyclically matched. If the matching is successful, the attribute is directly extracted. If the matching fails, the BERT model is used for attribute similarity judgment to obtain the most likely attribute value.

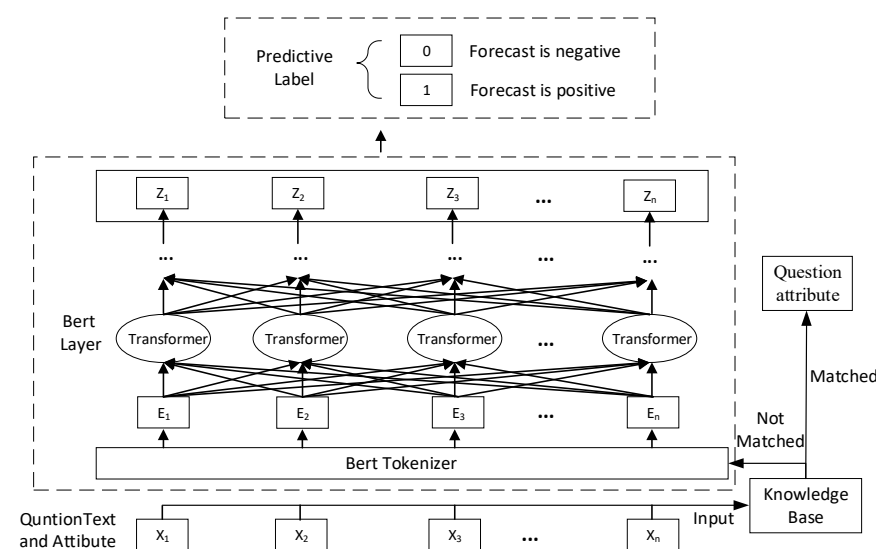
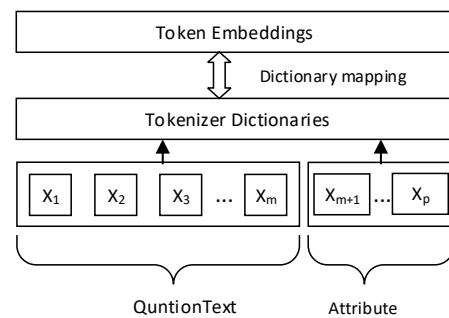


Figure 3. Question-type judgment based on the background knowledge base and BERT model.

To improve the accuracy of problem attribute similarity judgment, the BERT model was used for text feature extraction for input text. In this model, tokenizer processing was first performed [36], and the network architecture is shown in Figure 4. The main approach is to use a dictionary to embed tokens into the text, facilitating further feature extraction in the future.



**Figure 4.** Token embedding.

Token embedding represents the feature values of a single word itself at the word level; however, for text feature extraction, the order of words in the text is also crucial [37]. In the process of processing the BERT model, three types of features—namely, token features, position features, and segmented features [38]—will be fused. The calculation formula for the three feature vectors, where  $W$  is the number of characters in the sentence and  $d_m$  represents the overall dimension of the word vector in the question sentence, is as follows:

$$X_{Token} = R_1^{w \times d_m} \quad (9)$$

$$X_{Position} = R_2^{w \times d_m} \quad (10)$$

$$X_{Segment} = R_3^{w \times d_m} \quad (11)$$

After integrating the above three features, the BERT model further encodes the features through a multi-head attention mechanism, as shown in Formula (12). The principle is to multiply the random matrix  $W$  with the vector of the embedding layer to obtain three vectors: Query (Q), Key (K), and Value (V) [39].

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

To prevent the effect of an iteration being too bad, after the multi-head attention calculation, the model carries out residual connection and layer normalization operations and converts the input vector into data with a mean value of 0 and a variance of 1. The normalization formula is as follows, where  $E[x]$  is the mean value of the sample and  $\sigma^2[x]$  is the variance of the sample:

$$y = \frac{x_i - E[x]}{\sqrt{\sigma^2[x] + \varepsilon}} * r + \beta \quad (13)$$

### 3.3. Answer Generation

In the answer generation process, the method based on semantic analysis is mainly used for semantic analysis of questions; converting natural language questions into database semantic query statements; and, finally, generating natural language answers using background knowledge. The specific algorithm is shown in Table 1. From this, it can be seen that the knowledge-graph-based question-and-answer method is relatively traceable compared to large models because the answer is the answer in the knowledge base, and

the traceability method is not the focus of this paper; thus, it will not be introduced too much here. The focus here is on how to convert natural language questions into database query statements:

As shown in Table 1, the main conversion includes 6 steps. Firstly, the user inputs the question text they want to query (step 1 in Table 1); then, they identify the question entity using the entity recognition model (step 2 in Table 1). The specific principle is based on the question entity recognition model in this paper.

Afterward, based on the question entity, query all relevant triples in the database (steps 3 and 4 in Table 1). It should be specified that a triplet is represented as a row in a processed knowledge database, with each row having three fields, namely, the header entity, relationship, and tail entity. Therefore, it is relatively simple to query all related triplet rows based on the header entity.

Then, sort all associated triples and questions into the question discrimination model to obtain the best triplet option (step 5 in Table 1) and refer to the question type discrimination model in this paper for specific methods. Finally, the answer is returned to the user (step 6 in Table 1).

**Table 1.** Algorithm flow.

Algorithm Name		Answer Generation
Input	Question Qtext	
Output	Answer C	
Algorithm steps	1: User input question Qtext.	
	2: Input Qtext into the question entity recognition model (BERT-BiLSTM-CRF) to obtain the question entity set $E \{e1, e2 \dots en\}$ .	
	3: Generate knowledge base query statement $S \{s1, s2 \dots sn\}$ according to entity set $E \{e1, e2 \dots en\}$ .	
	4: Retrieve the knowledge base query statement $S \{s1, s2 \dots sn\}$ in the knowledge base to obtain the associated triple set.	
	5: Input the question Qtext and associated triple $T$ into the attribute similarity judgment model (BERT) to determine the best triple $T_k = \langle A, B, C \rangle$ . A represents the head entity, B represents the relationship, and C represents the tail entity.	
	6: Output Answer C.	

## 4. Results

### 4.1. Experimental Data Set

The data set used in this paper is the public data set NLPCC2016-KBQA [40] provided by the NLPCC (The Conference on Natural Language Processing and Chinese Computing). The training set in this data set contains 14,609 question-and-answer pairs, and the test set contains 9870 question-and-answer pairs. The sample data of the question-and-answer pair are shown in Table 2.

**Table 2.** Q&A pair data sample.

No.	Interrogative Sentence	Triplet	Answer
1	What are the main crops in Quanshan Street?	Quanshan Street     main crops     Wheat, corn, sweet potatoes, peanuts	Wheat, corn, sweet potatoes, peanuts
2	How did Lion Stone get its name?	Lion stone     Reason for Naming     Mount resembling Manjushri Bodhisattva	Mount resembling Manjushri Bodhisattva

Before the training of the question entity recognition model, the questions in the public data set NLPCC2016-KBQA are labeled and preprocessed, and the BIO labeling method is used. The labeled data sample is shown in Table 3.

**Table 3.** Sample question annotation.

Question	How	did	Lion	Stone	get	its	name
Tagging	O	O	B	I	O	O	O

The background knowledge used in this paper is the knowledge base provided by NLPCC2016-KBQA, which contains 43,063,796 rows of triples, and the knowledge base sample is shown in Table 4.

**Table 4.** Knowledge base triplet example.

No.	Head Entity	Attribute Relationship	Tail Body
1	«The Dream of Red Mansion»	Leading actor	Qian Huili; Dan Yangping; Chen Ying; Fang Yafen
2	«The Dream of Red Mansion»	Release time	2005
3	«The Dream of Red Mansion»	Production Company	China Film Group, Rongxinda, Hualu Baina

#### 4.2. Experimental Environment and Evaluation Index

In terms of the experimental environment setting, the experimental code is mainly based on the Python and PyTorch framework. The Python version used is 3.6, the PyTorch version is 1.7.1, and the transformer module version is 2.5.1. The learning rate of the question entity recognition model and the attribute similarity judgment model is  $5 \times 10^{-5}$ . The model optimizer is Adam [41].

Based on the mainstream experience of knowledge Q&A model evaluation, this paper selects Precision rate, Recall rate, and F1 value as the evaluation indicators for problem entity recognition and problem similarity judgment [42]. The calculation method is as follows:

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F_1 = \frac{2 * P * R}{P + R} \quad (16)$$

In the above formula,  $TP$  represents the correct number of predictions,  $P$  represents the precision rate of the experiment, and  $R$  represents the recall rate of the experiment;  $FP$  refers to the number of triples predicted incorrectly,  $TP + FP$  refers to the total number of triples predicted,  $FN$  refers to the number of triples predicted incorrectly, and  $TP + FN$  refers to the number of triples marked.

In addition, to ensure the reliability of the experimental results, this paper conducted multiple experiments (five times) on the knowledge question-and-answer task, and the final results were taken as the average.

#### 4.3. Comparative Analysis of Results

This paper uses NLPCC2016-KBQA to carry out experimental verification on the two key models of the question entity recognition model and attribute similarity judgment model in the method model. In question entity recognition, the public data set NLPCC2016-KBQA was used to compare the traditional BiLSTM-CRF entity recognition model with

the proposed BERT–BiLSTM–CRF question entity recognition model and its variants, including introducing a variant of the BERT model, Roberta [43], into the baseline model for comparison [44]. The experimental results are shown in Table 5.

**Table 5.** Overall evaluation results of question entity recognition.

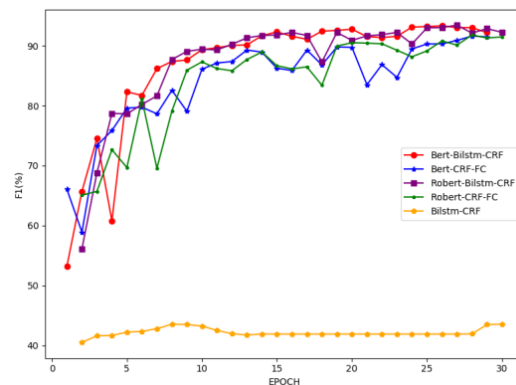
Model	Average Precision	Average Recall Rate	Average F1 Value	Best Epoch
CRF	29.5	81.5	41.6	12
BiLSTM–CRF	31.8	83.4	43.2	8
BAT–KBQA [45]	–	–	87.7	–
BERT–CRF	94.6	95.6	95.1	91
BERT–BiLSTM–CRF	95.1	96.0	95.5	92
BERT–BiGRU–CRF	94.3	95.0	94.5	90
RoBERTa–CRF	96.4	94.2	95.3	94
RoBERTa–BiLSTM–CRF	93.9	96.9	95.4	98

The result of question entity recognition for a single tag is shown in Table 6.

**Table 6.** Itemized evaluation results of question entity identification label.

Model	Label B Prediction Precision Rate	Label B Prediction Recall Rate	Label B Prediction F1	Label I Prediction Precision Rate	Label I Prediction Recall Rate	Label I Prediction F1
CRF	23.2	4.2	8.3	34.1	97.3	48.2
BiLSTM–CRF	24.1	5.7	9.3	33.3	98.4	49.8
BERT–CRF	95.6	93.6	94.6	94.4	96.0	95.1
BERT–BiLSTM–CRF	97.3	91.4	94.3	94.6	97.0	95.7
BERT–BiGRU–CRF	96.5	91.2	93.1	93.2	95.0	94.3
RoBERTa–CRF	96.7	93.4	95.3	96.3	94.3	95.3
RoBERTa–BiLSTM–CRF	93.4	95.0	94.2	94.0	97.3	95.6

The F1 value transformation of different models at various stages is shown in Figure 5.



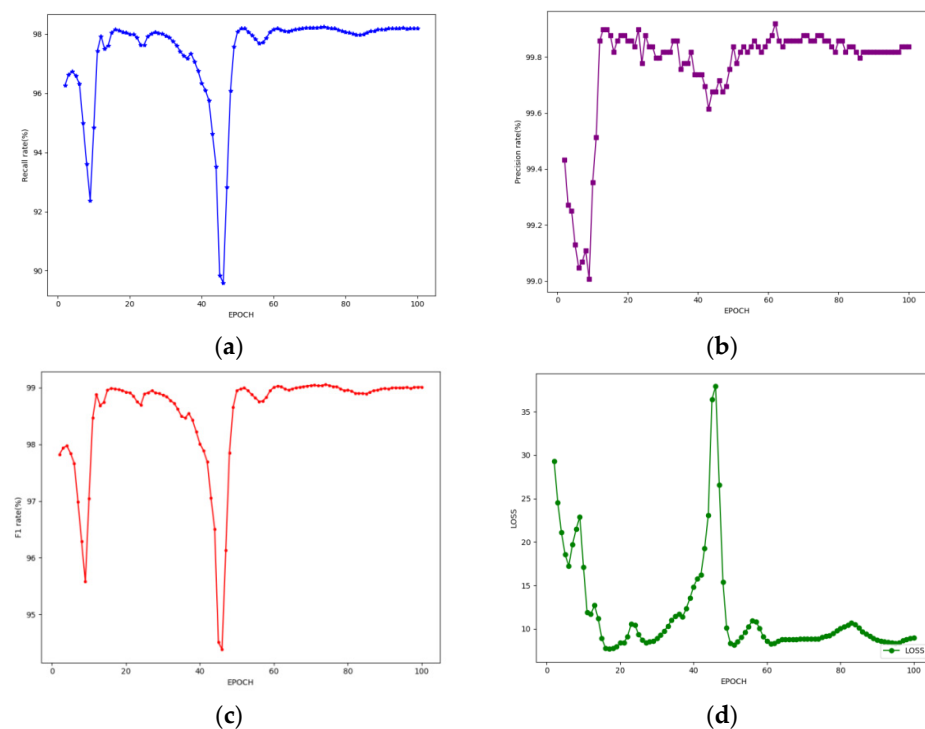
**Figure 5.** Experimental results of question entity recognition.

From the above experimental results, we can see that the F1 value of the BERT–BiLSTM–CRF model for the question entity recognition task reached 95.5%, slightly higher than its similar variants and about 54% higher than the traditional BiLSTM–CRF entity recognition model. As for the independent evaluation indicators of B and I labels, the BERT–BiLSTM–CRF model and RoBERTa–BiLSTM–CRF model perform better, each with advantages and disadvantages. In terms of the overall effect, the BERT–BiLSTM–CRF model has a slightly better effect than the RoBERTa–BiLSTM–CRF model. This also proves the effectiveness of the method proposed in this paper. In the public data set NLPCC2016–KBQA, the experimental results of the BERT model for the question category subtask are shown in Table 7.

**Table 7.** Triad attribute similarity judgment and evaluation index.

Epochs	5	50	100
Precision rate	96.396	97.741	98.396
Recall rate	99.473	99.919	99.878
F1 Value	97.910	98.818	99.131
Test set LOSS	26.967	17.191	8.592

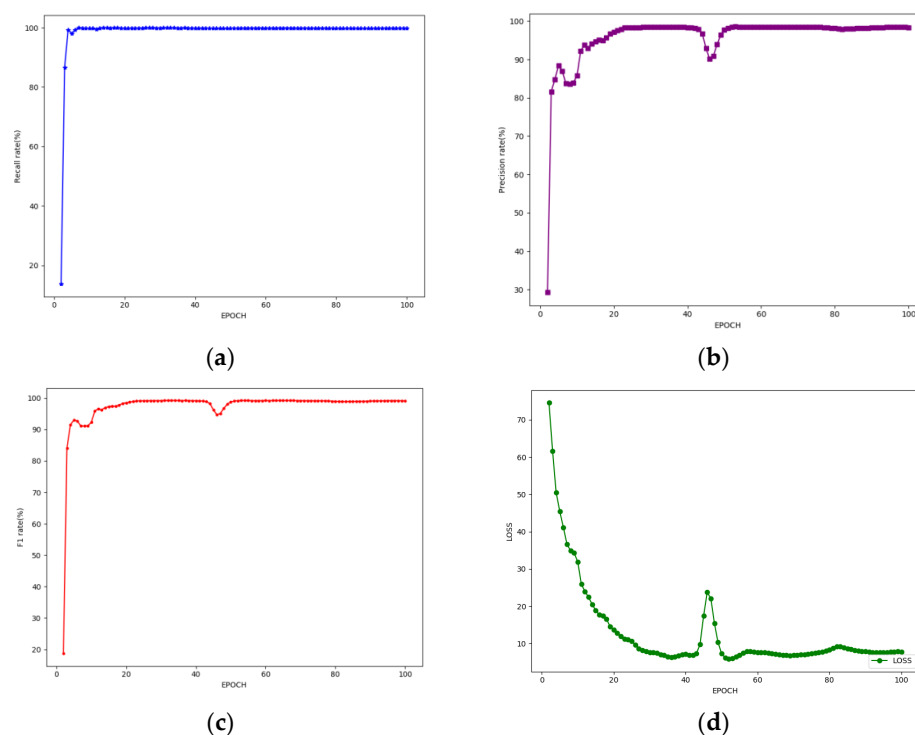
It can be seen in Figure 6 that the accuracy effect and the loss value will shake at the 10th and 43rd epochs. To explore whether this phenomenon is accidental, this paper conducted a second experiment, and the experimental results are shown in Figure 7.



**Figure 6.** Experimental results of attribute similarity judgment based on BERT model (1). (a) The change process of Recall Rate; (b) The change process of Precision Rate; (c) The change process of F1 value; (d) The change process of Loss value.

From the above results, it can be seen that the method may have jitteriness at a specific stage, which may be caused by the experimental data. However, the small jitter amplitude does not affect the overall experimental effect of the method. From the above experimental results, in the judgment of question type, the overall accuracy of the BERT model can reach 98.39%. The lowest point is accurate at more than 85%, and the overall effect is better in other cases.

From the above experimental results, it can be seen that for the question entity recognition task, the proposed model significantly outperforms those without using a BERT encoder regarding F1. Compared with its variant model, it can also achieve the best F1 value; according to the best epoch in Table 5, it can be seen that compared with other BERT variants, its performance is also very competitive. For the question-type judgment task, the proposed model can achieve an F1 value of 98.39%. Figures 6 and 7 show that the model's performance is quite stable with regard to the evaluation metrics. Overall, the model is highly competent in knowledge question answering.



**Figure 7.** Experimental results of attribute similarity judgment based on BERT model (2). (a) The change process of Recall Rate; (b) The change process of Precision Rate; (c) The change process of F1 value; (d) The change process of Loss value.

## 5. Error Analysis

For the task of question entity recognition, it can be seen in Figure 5 that the F1 values of each model fluctuate less and grow steadily. Table 6 shows that the model performs smoothly on each label and there is no imbalance. However, other models (such as the CRF model and the BiLSTM–CRF model) exhibit significant differences in performance between the B label and the I label. We believe that this may be due to the feature encoder of the model, as other models using BERT as the feature encoder do not have this problem.

We conducted two experiments on the question-type judgment task and found that F1 jitter occurred during the 40th to 50th epochs. However, the magnitude of the jitter was different between the two experiments. We believe that it is an overfitting jitter caused by the lack of an experimental corpus. However, this jitter will not affect the final experimental results; so, we will expand the corpus in future research to further investigate this issue.

## 6. Conclusions

For knowledge question-and-answer tasks, this paper makes a comprehensive comparison and analysis of the existing LLM-based methods, rule-based question-and-answer methods, information-based question-and-answer methods, and semantic-analysis-based question-and-answer methods. Based on the research of semantic-analysis-based question-and-answer methods, a question-and-answer method based on the BBCB (BERT–BiLSTM–CRF–BERT) model and background knowledge is proposed, and experimental verification is carried out on the open data set. In the follow-up work, we hope to continue to improve the training effect of the model, especially considering the auxiliary method of adding a high-precision dictionary to improve the overall accuracy of the question and answer.

**Author Contributions:** Y.L.: conceptualization, writing—original draft, writing—reviewing and editing, methodology, validation, formal analysis, data curation. H.Z.: supervision, writing—review and editing. T.Z. and J.W.: supervision, formal analysis. W.D.: writing—review and editing. All authors contributed to the paper and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data sets generated during the current study are not publicly available as the data also form part of an ongoing study; however, they are available from the corresponding author on reasonable request.

**Acknowledgments:** The work described in this study was conducted at the School of Information Communication, National University of Defense Technology. The authors would like to thank Zuo and Fu for their support and encouragement.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liu, H.; Zhang, Y.; Li, Y.; Kong, X. Review on emotion recognition based on electroencephalography. *Front. Comput. Neurosci.* **2021**, *15*, 84. [CrossRef]
2. Othman, N.; Faiz, R.; Smaïli, K. Learning english and arabic question similarity with siamese neural networks in community question answering services. *Data Knowl. Eng.* **2022**, *138*, 101962. [CrossRef]
3. Maheshwari, G.; Trivedi, P.; Lukovnikov, D.; Chakraborty, N.; Fischer, A.; Lehmann, J. Learning to rank query graphs for complex question answering over knowledge graphs. In *The Semantic Web—ISWC 2019, Proceedings of the 18th International Semantic Web Conference, Auckland, New Zealand, 26–30 October 2019; Part I*; Springer International Publishing: Cham, Switzerland, 2019; pp. 487–504.
4. Jabalameli, M.; Nematbakhsh, M.; Ramezani, R. Denoising distant supervision for ontology lexicalization using semantic similarity measures. *Expert Syst. Appl.* **2021**, *177*, 114922. [CrossRef]
5. Jabalameli, M.; Nematbakhsh, M.; Zaeri, A. Ontology-lexicon-based question answering over linked data. *ETRI J.* **2020**, *42*, 239–246. [CrossRef]
6. Lin, L.; Liu, J.; Zhang, X.; Liang, X. Automatic translation of spoken English based on improved machine learning algorithm. *J. Intell. Fuzzy Syst.* **2021**, *40*, 2385–2395. [CrossRef]
7. Jiang, Y.; Yang, G.; Li, H.; Zhang, T. Knowledge driven approach for smart bridge maintenance using big data mining. *Autom. Constr.* **2023**, *146*, 104673. [CrossRef]
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
9. Wang, R.Z.; Ling, Z.H.; Hu, Y. Knowledge base question answering with attentive pooling for question representation. *IEEE Access* **2019**, *7*, 46773–46784. [CrossRef]
10. Bi, X.; Nie, H.; Zhang, G.; Hu, L.; Ma, Y.; Zhao, X.; Yuan, Y.; Wang, G. Boosting question answering over knowledge graph with reward integration and policy evaluation under weak supervision. *Inf. Process. Manag.* **2023**, *60*, 103242. [CrossRef]
11. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
12. Green, B.F., Jr.; Wolf, A.K.; Chomsky, C.; Laughery, K. Baseball: An automatic question-answerer. In *Proceedings of the Western Joint IRE-AIEE-ACM Computer Conference*, Los Angeles, CA, USA, 9–11 May 1961; pp. 219–224.
13. Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM* **1966**, *9*, 36–45. [CrossRef]
14. Antoniou, C.; Bassiliades, N. A survey on semantic question answering systems. *Knowl. Eng. Rev.* **2022**, *37*, e2. [CrossRef]
15. Lan, Y.; He, G.; Jiang, J.; Jiang, J.; Zhao, W.X.; Wen, J.-R. Complex knowledge base question answering: A survey. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 11196–11215. [CrossRef]
16. Zhong, B.; He, W.; Huang, Z.; Love, P.E.; Tang, J.; Luo, H. A building regulation question answering system: A deep learning methodology. *Adv. Eng. Inform.* **2020**, *46*, 101195. [CrossRef]
17. Mahmud, T.; Hasan, K.A.; Ahmed, M.; Chak, T.H.C. A rule based approach for NLP based query processing. In *Proceedings of the 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT)*, Khulna, Bangladesh, 10–12 December 2015; pp. 78–82.
18. Woods, W.A. Progress in natural language understanding: An application to lunargeology. In *Proceedings of the National Computer conference and Exposition*, New York, NY, USA, 4–8 June 1973; pp. 441–450.
19. Liu, Z.J.; Wang, X.L.; Chen, Q.C.; Zhang, Y.Y.; Xiang, Y. A Chinese question answering system based on web search. In *Proceedings of the 2014 International Conference on Machine Learning and Cybernetics*, Malaga, Spain, 13–16 July 2014; Volume 2, pp. 816–820.
20. Riloff, E.; Thelen, M. A rule-based question answering system for reading comprehension tests. In *ANLP-NAACL 2000 Workshop: Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems*; 2000. Available online: <https://aclanthology.org/W00-0603.pdf> (accessed on 4 May 2020).
21. Dong, L.; Wei, F.; Zhou, M.; Xu, K. Question answering over freebase with multi-column convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, China, 26–31 July 2015; pp. 260–269.

22. Bordes, A.; Weston, J.; Usunier, N. Open question answering with weakly supervised embedding models. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; ACM: Nancy, France, 2014; pp. 165–180.
23. Zhang, Y.; Liu, K.; He, S.; Ji, G.; Liu, Z.; Wu, H.; Zhao, J. Question answering over knowledge base with neural attention combining global knowledge information. *arXiv* **2016**, arXiv:1606.00979.
24. Jiajie, D.; Kang, X.; Heng, Y.; Zhou, X.; Zhang, M. Data enhancement method for question answering field. *J. Peking Univ. Nat. Sci. Ed.* **2022**, *58*, 54–60.
25. Peng, Y.; Li, X.; Hu, S.; Liu, X.; Qian, W. Three-stage question answering model based on BERT. *J. Comput. Appl.* **2022**, *42*, 64–70.
26. Li, J.; Luo, Z.; Huang, H.; Ding, J. Towards Knowledge-Based Tourism Chinese Question Answering System. *Mathematics* **2022**, *10*, 664. [\[CrossRef\]](#)
27. Shamsabadi, A.S.; Ramezani, R.; Farsani, H.K.; Nematbakhsh, M. Direct relation detection for knowledge-based question answering. *Expert Syst. Appl.* **2023**, *211*, 118678. [\[CrossRef\]](#)
28. Antaki, F.; Touma, S.; Milad, D.; El-Khoury, J.; Duval, R. Evaluating the Performance of ChatGPT in Ophthalmology: An Analysis of its Successes and Shortcomings. *Ophthalmol. Sci.* **2023**, *3*, 100324. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Zhang, P.J.; Jia, Y.H.; Chen, W.L. Chinese knowledge based question answering based on multi-feature entity disambiguation. *Comput. Eng.* **2022**, *48*, 47–54.
30. Yang, Y.; Yih, W.; Meek, C. Wikiqa: A challenge dataset for open-domain question answering. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 17–21 September 2015; pp. 2013–2018.
31. Singhal, A. Introducing the Knowledge Graph: Things, not Strings. Official Google Blog. **2012**, *5*, 16. Available online: <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/> (accessed on 16 May 2012).
32. Zhang, J.; Zhang, L.; Hui, B.; Tian, L. Improving complex knowledge base question answering via structural information learning. *Knowl. Based Syst.* **2022**, *242*, 108252. [\[CrossRef\]](#)
33. Duan, N. Overview of the nlpcc-iccpol 2016 shared task: Open domain Chinese question answering. In *Natural Language Understanding and Intelligent Applications, Proceedings of the 5th CCF Conference on Natural Language Processing and Chinese Computing, NLPCC 2016, and 24th International Conference on Computer Processing of Oriental Languages, ICCPOL 2016, Kunming, China, 2–6 December 2016*; Springer International Publishing: Cham, Switzerland, 2016; pp. 942–948.
34. Zhong, Q.; Tang, Y. An attention-based BiLSTM-CRF for Chinese named entity recognition. In *Proceedings of the 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, Chengdu, China, 10–13 April 2020; pp. 550–555.
35. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Liu, Y.; Wen, F.; Zong, T.; Li, T. Research on joint extraction method of entity and relation triples based on hierarchical cascade labeling. *IEEE Access* **2022**, *11*, 978. [\[CrossRef\]](#)
37. Fan, C.; Chen, W.; Wu, Y. Knowledge base question answering via path matching. *Knowl. Based Syst.* **2022**, *256*, 109857. [\[CrossRef\]](#)
38. Li, R.; Wang, L.; Jiang, Z.; Liu, D.; Zhao, M.; Lu, X. Incremental BERT with commonsense representations for multi-choice reading comprehension. *Multimed. Tools Appl.* **2021**, *80*, 32311–32333. [\[CrossRef\]](#)
39. Tian, D.; Li, M.; Ren, Q.; Zhang, X.; Han, S.; Shen, Y. Intelligent question answering method for construction safety hazard knowledge based on deep semantic mining. *Autom. Constr.* **2023**, *145*, 104670. [\[CrossRef\]](#)
40. Wu, W.; Zhu, Z.; Qi, J.; Wang, W.; Zhang, G.; Liu, P. A dynamic graph expansion network for multi-hop knowledge base question answering. *Neurocomputing* **2023**, *515*, 37–47. [\[CrossRef\]](#)
41. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
42. Wang, Z.; Huang, Z.; Gao, J. Chinese text classification method based on BERT word embedding. In *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*, Beijing, China, 10–12 May 2020; pp. 66–71.
43. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
44. Liu, Y.; Zuo, Q.; Wang, X.; Zong, T. Entity relationship extraction based on a multi-neural network cooperation model. *Appl. Sci.* **2023**, *13*, 6812. [\[CrossRef\]](#)
45. Duan, K.; Du, S.; Zhang, Y.; Lin, Y.; Wu, H.; Zhang, Q. Enhancement of Question Answering System Accuracy via Transfer Learning and BERT. *Appl. Sci.* **2022**, *12*, 11522. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.