Using BERT-CRF to Build an Automatic English Terminology Annotation Engine for Higher Vocational Courses

Jingjing Zhu[1], Xiaoran Li[1], Tao Song[2], Na Hu[1], Liping Zhang[1]
[1]Department of Basic Curriculum Teaching, Shandong Water Conservancy Vocational College, Rizhao 276826, Shandong, China
[2]Rizhao Water Supply Company Limited, Rizhao 276800, Shandong, China

Corresponding author: Tao Song (e-mail: st1437zjj@163.com)

ABSTRACT In higher vocational education, professional English terminology emphasizes more precise semantic expression in specific professional fields than general academic English. This demand for terminological precision is critical within technical disciplines where compound and nested terms carry strict professional meanings. Existing annotation methods, relying on general corpus training, struggle to accurately handle the compound and nested terminology common in higher vocational courses. To address this, this paper introduces the TermTag-BERTCRF model, an English term automatic tagging engine. The model enhances performance by pre-training BERT on a specific vocational corpus and using a BiLSTM-CRF structure to optimize the recognition of compound and nested term boundaries. It integrates multi-level manual features and employs a dual-dimensional labeling system for multi-granularity annotation. The deployed lightweight model offers efficient batch processing and real-time labeling via API and visualization. Experiments show the TermTag-BERTCRF model outperforms the baseline, achieving an F1 score of 91.0% overall, and high accuracy for complex 3-word (85.4%) and 4-word (78.7%) terms. With fast response (128.6 ms) and strong stability under noise, this model provides effective and promising support for the automatic annotation and knowledge extraction of English terms in higher vocational education. This technology can benefit industry-specific training by precisely identifying and extracting complex technical terminology, thereby improving the efficiency of creating accurate domain-specific educational resources. The same annotation logic can support terminology organization in courses involving antennas, waveguides and impedance matching.

INDEX TERMS english terminology annotation,BERT-CRF model,vocational engineering education,technical terminology,antenna terminology

I. INTRODUCTION

WITH the rapid development of vocational education [1-3], the application of English terminology in higher vocational courses is becoming more and more extensive, a trend profoundly affecting technical disciplines . As a result, precise English terminology has become essential for students in programs to understand complex machinery operations, and international trade doc-uments [4-6]. However, due to the high level of specialization and complexity of higher vocational courses [7-9], existing annotation methods still face many challenges in the automatic annotation of English terms in this context [10-12]. On the one hand, traditional natural language processing technology usually relies on general corpus training and lacks indepth understanding of specific terms [13-15]. On the other hand, English terms in higher vocational courses often contain compound structures and nested forms [16,17], which further in-creases the difficulty of automatic annotation.

To address the above challenges, some studies have attempted to introduce deep learning technology into the terminology annotation process, but there are still some defects. For example, the deep learning model constructed by Li et al. [18] could effectively and automatically extract professional terms from the original text and classify the extracted results according to the original terminology database. However, they did not consider the problems of low-frequency terms and nested structures, and lacked support for labeling systems and fine-grained classification. Wang et al. [19] systematically explored the feature extraction methods of deep learning models in natural language processing and applied them to the task of learning English terminology, which improved the model's ability to understand language structure and semantic information. However, they did not optimize terminology in specific fields and lacked a fine-grained recognition mechanism for low-frequency words and complex structures. Guo et al. [20] applied deep learning models to the task of annotating

English texts, but did not further identify and classify terms and lacked an evaluation of the model's cross-domain generalization capabilities. Therefore, there is an urgent need for a new method that can balance term recognition accuracy, reasoning efficiency, and cross-domain adaptability.

Based on this, this paper proposes TermTag-BERTCRF, an English terminology automatic annotation engine for higher vocational education, built on a BERT–CRF framework (BERT: Bidirectional Encoder Representations from Transformers; CRF: Conditional Random Field) [21,22]. The model achieves accurate recognition and efficient annotation of English terms in higher vocational courses through multiple key steps: first, BERT [23- 25] is pre-trained twice on higher vocational English corpus to enhance its ability to understand professional terms. Second, a fine-grained label decoding module based on CRF [26-28] is designed, combined with Bidirectional Long Short-Term Memory (BiLSTM) [29-31] to extract deep sequence features. The label transfer matrix is modeled through the CRF layer to optimize the recognition of compound and nested term boundaries. Furthermore, multi-level manual features such as part of speech, word frequency, position and first letter capitalization are integrated to improve the discriminability of input representation. At the same time, a multidimensional label system based on the Begin, Inside, Outside, End, Single (BIOES) format and the subjecttype dual dimensions is constructed to support fine-grained and nested structure annotation. Finally, a weighted cross entropy loss function and a label transfer penalty mechanism are introduced to alleviate the problem of low-frequency term recognition. Existing automatic terminology annotation methods face two key bottlenecks when dealing with English terminology in the context of higher vocational education: the first is the challenge of recognizing nested structures. Higher vocational courses commonly contain multi-layered nested terms such as "adaptive finite element method" or "real-time fault diagnosis system", whose internal components can themselves constitute legitimate sub-terms (e.g., "finite element method"). Traditional linear annotation systems (BIO) struggle to accurately model such structures, easily leading to boundary misalignment or omissions. The second is the difficulty of recognizing low-frequency terms. Due to the highly specialized nature of professional fields, many terms (e.g., "sericin recovery process") appear with extremely low frequency in the corpus, making models susceptible to interference from high-frequency general terms, resulting in semantic shifts or misjudgments. Although deep learning models perform excellently in general named entity recognition, when directly transferred to higher vocational terminology scenarios, they often lack domain adaptation mechanisms and fine-grained annotation support, making it difficult to balance boundary accuracy and sparse term recall. Therefore, a terminology annotation framework that integrates domain knowledge, supports nested structure modeling, and possesses low-frequency term enhancement capabilities is urgently needed.

The experimental results of this paper show that the TermTag-BERTCRF model outperforms the baseline model in many indicators, with accuracy, recall, and F1 scores reaching 92.4%, 89.7%, and 91.0%, respectively. In addition, the model also has efficient reasoning ability and stability, with a single document response time of 128.6 ms and a memory usage of 379.4 MB, and performs well in cross-course generalization ability evaluation. In summary, the TermTag-BERTCRF model proposed in this study provides an effective solution for the automatic annotation of English terms in higher vocational education. This offers a significant practical advantage for specialized training, where the model can accurately tag complex terms.

II. ALGORITHM DESIGN

This study focuses on improving the domain adaptability and fine-grained annotation capability of English terminology recognition in higher vocational courses. The system is designed from multiple perspectives, including model architecture, feature fusion, label construction, optimization objectives, and deployment. Firstly, BERT is pre-trained twice on vocational English corpus to enhance its ability to understand professional terminology. Secondly, the CRF decoding module is used in combination with BiLSTM to extract deep sequence features and optimize the recognition of compound and nested term boundaries. Furthermore, multilevel manual features such as part of speech, word frequency, position and capitalization of the first letter are integrated to improve the discriminability of input representation. A multidimensional labeling system based on the BIOES format and the discipline-type dual dimension is constructed to support fine-grained and nested structure labeling; a weighted cross entropy loss function and a label transfer penalty mechanism are introduced to alleviate the sample imbalance problem and improve the recognition rate of low-frequency terms; finally, a lightweight model is deployed, and the Application Programming Interface (API) service and the front-end visualization system are integrated to achieve efficient batch processing and real-time labeling functions. The overall algorithm design framework of this study is shown in Figure 1.

A. DOMAIN-ADAPTED BERT MODEL PRE-TRAINING

To improve the domain adaptability of the BERT model in the task of English terminology recognition in higher vocational courses, this paper designs a set of BERT model secondary pre-training processes for professional contexts. In the corpus construction stage, this paper collects English text resources related to higher vocational education from multiple sources to ensure that BERT has the ability to generalize across courses. In the data preprocessing stage, regular expressions are used to perform structural cleaning of the original text: Hypertext Markup Language (HTML) tags, special symbols, and low-quality sentence fragments are removed, and Spacy is called to complete fine-grained part-of-speech tagging and dependency parsing. Finally,

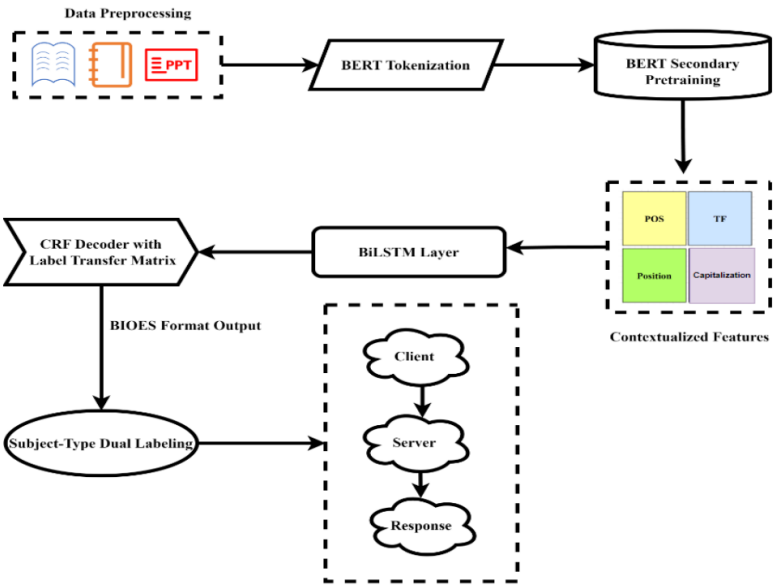

FIGURE 1. Algorithm design framework

a token sequence in a unified format is output as input for subsequent training; at the model architecture level, incremental secondary pre-training is performed on the basis of the existing BERT model. This strategy aims to retain the ability to understand general language while strengthening the model's ability to understand and represent terms in the field of higher vocational education.

Assuming that the original BERT parameters are θ_{base}, the parameter update process under the new training objective can be expressed as:

$$\theta_{\text{new}} = \arg\max_{\theta} \sum_{(x^{(i)},y^{(i)})\in D} \log P\left(y^{(i)} \mid x^{(i)};\theta\right) \quad (1)$$

In formula 1, D represents the pre-training corpus; $x^{(i)}$ is the input sequence; $y^{(i)}$ is the real token at the masked position.

The loss function uses the standard cross entropy form:

$$L_{\text{MLM}} = -\frac{1}{N}\sum_{t=1}^{T}\sum_{k=1}^{V} y_{t,k}\log p_{t,k} \quad (2)$$

In formula 2, T is the sequence length; V is the vocabulary size; $y_{t,k}$ is the one-hot encoding of the k-th item in the vocabulary of the real token at the t-th position; $p_{t,k}$ is the probability distribution predicted by the model.

The two pre-training stages referred to in this paper do not correspond to two independent training tasks; rather, they are two stages of domain-adaptive pre-training performed sequentially on top of the general

BERT-base model: the first stage uses large-scale unlabeled vocational college English text (approximately 5 million words) for Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) tasks; the second stage, based on the model weights from the first stage, uses refined teaching documents (such as textbooks and lab manuals, approximately 1.2 million words) for further MLM-only pre-training to enhance the semantic density of terms. Both stages share the same vocabulary, with a total of 120 K training steps (80 K in the first stage and 40 K in the second).

B. DESIGN OF FINE-GRAINED LABEL DECODING MODULE BASED ON CRF

To improve the recognition accuracy of the model for compound structures and nested terms in English terms in higher vocational courses, this paper constructs a fine-grained label decoding module based on CRF. This module introduces a bidirectional long short-term memory network based on BERT output to further extract sequence features and model the transfer constraint relationship between labels through the CRF layer to optimize term boundary recognition and label consistency.

Specifically, the input sequence is the context vector representation $H = [h_1, h_2, \ldots, h_T] \in R^{T\times d}$ encoded by BERT, where T is the sequence length, and d is the hidden-state dimension.

While BERT possesses powerful contextual modeling capabilities, its output vector still exhibits ambiguity at local sequence boundaries (such as the start and end positions of multi-word terms) when facing highly structured compound and nested term recognition tasks in higher vocational education. The BiLSTM layer is not redundant here, but rather serves as a lightweight sequence smoothing and boundary strengthening mechanism specifically designed to capture word order dependencies and adjacency constraints within terms. In experiments, the baseline model using only BERT-CRF achieves an F1 score of 81.3% for terms with three or more words, which improves to 85.4% after adding

BiLSTM, validating its crucial role in fine-grained boundary recognition. Furthermore, the BiLSTM hidden unit dimension in this model is set to 128 (significantly smaller than BERT's 768 dimensions) and optimized using TorchScript static graphs, adding only about 9 milliseconds of latency during inference while still satisfying the requirements for lightweight deployment. This layer is jointly fine-tuned end-to-end with the BERT encoder to ensure that the feature representation is highly aligned with the task objective.

Subsequently, H is sent to the BiLSTM layer to extract deeper sequence dependencies:

$$\overrightarrow{h_t} = \mathrm{LSTM}_{\mathrm{fwd}}\left(h_t, \overrightarrow{h_{t-1}}\right) \tag{3}$$

$$\overleftarrow{h_t} = \mathrm{LSTM}_{\mathrm{bwd}}\left(h_t, \overleftarrow{h_{t+1}}\right) \tag{4}$$

$$h'_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{5}$$

In formula 5, h'_t is the fusion representation of the t-th token in the BiLSTM layer, with a dimension of $2d$. Finally, the enhanced sequence representation matrix $H' = [h'_1, h'_2, \ldots, h'_T] \in R^{T\times 2d}$ is obtained. Next, H' is input to the CRF layer for label decoding. Let the label set be $Y = \{y_1, y_2, \ldots, y_K\}$, where K is the total number of labels, using the BIOES annotation format.

CRF models the state transition probability between labels by constructing the label transfer matrix $T \in R^{K\times K}$:

$$\log P(y \mid X) = \sum_{t=1}^{T} W^{\top}\phi(y_t, h'_t) + \sum_{t=2}^{T} T_{y_{t-1},y_t} - \log Z(X) \tag{6}$$

In Formula 6, $\phi(y_t, h'_t)$ is the emission feature function; W denotes learnable weight; T_{y_{t-1},y_t} is the transition score from y_{t-1} to y_t; and $Z(X)$ is the normalization factor, which is defined as follows:

$$Z(X) = \sum_{y\in Y^T} \exp\left(\sum_{t=1}^{T} W^{\top}\phi(y_t, h'_t) + \sum_{t=2}^{T} T_{y_{t-1},y_t}\right) \tag{7}$$

At the implementation level, this paper uses the PyTorch-CRF library under the PyTorch framework to implement the CRF layer, which uses the Viterbi algorithm to search for the optimal path to ensure that the decoding result meets the label transfer constraint. At the same time, to improve the efficiency of model reasoning, the CRF layer is lightweight, and the transfer matrix is restricted to only allow conversions between legal label combinations.

C. MULTI-LEVEL FEATURE FUSION STRATEGY

To enhance the model's ability to recognize English terms in higher vocational courses and improve the diversity and discriminability of input representation, this paper designs a multi-level manual feature fusion mechanism. Based on BERT output, this mechanism introduces four types of structural linguistic features, including part of speech (POS), term frequency (TF), positional feature, and capitalization, and achieves end-to-end fusion through vectorized encoding and concatenation operations.

For each token t, the following four types of features are extracted: SpaCy is used for automatic annotation to obtain the part-of-speech tag $p_t \in P$ of the current token, where P is a predefined part-of-speech set; the frequency of each token in the entire training corpus is counted $f_t = \log(1 + \mathrm{count}(w_t))$ to alleviate the problem of high-frequency words dominating the model; the relative position of the word in the sentence is calculated $\mathrm{pos}_t = t/T$, where T is the total length of the sentence; if the first letter of the word is capitalized, it is marked as 1, otherwise it is 0, recorded as $c_t \in \{0, 1\}$. Among the above features, discrete features (such as POS, Capitalization) are mapped to high-dimensional space using one-hot encoding:

$$e_t^p = \mathrm{OneHot}(p_t) \tag{8}$$

$$e_t^c = \mathrm{OneHot}(c_t) \tag{9}$$

In formula 8 and formula 9, $e_t^p \in R^{|P|}$, $e_t^c \in R^2$.

Continuous features (such as TF and Positional Feature) are directly used as numerical features after normalization:

$$\tilde{f}_t = \frac{f_t}{\max(f)} \tag{10}$$

$$\widetilde{\mathrm{pos}}_t = pos_t \tag{11}$$

All feature vectors are uniformly mapped to a low-dimensional dense representation space and concatenated with the BERT output. Assuming that the context representation of the t-th token in the BERT output layer is $h_t \in R^{d_{\mathrm{BERT}}}$, the fused feature representation is:

$$x_t = \left[h_t; E_p e_t^p; E_c e_t^c; \tilde{f}_t; \widetilde{\mathrm{pos}}_t\right] \tag{12}$$

In formula 12, $E_p \in R^{d_p\times|P|}$ and $E_c \in R^{d_c\times 2}$ are the embedding matrices of POS and Capitalization, respectively; d_p and d_c are the corresponding embedding dimensions. The concatenated feature vector $x_t \in R^{d_{\mathrm{total}}}$, where $d_{\mathrm{total}} = d_{\mathrm{BERT}} + d_p + d_c + 2$. The fused feature sequence $X = [x_1, \ldots, x_T]$ is input into a two-layer fully connected neural network for nonlinear transformation:

$$z_t = \mathrm{ReLU}(W_1 x_t + b_1) \tag{13}$$

$$h'_t = \mathrm{ReLU}(W_2 z_t + b_2) \tag{14}$$

In formula 13 and formula 14, $W_1 \in R^{d_h\times d_{\mathrm{total}}}$; $W_2 \in R^{d_h\times d_h}$; d_h is the hidden layer dimension. The final output sequence representation $H' = [h'_1, \ldots, h'_T] \in R^{T\times d_h}$ is input to the CRF layer to complete label decoding. The

fused feature vector x_t is first input into a two-layer fully connected network for nonlinear compression (Formulas 13–14), and its output z_t is used as the input to the BiLSTM layer. In other words, the concatenation of the handcrafted features and the BERT representation occurs before the BiLSTM, rather than being directly fed into the CRF. This design allows the BiLSTM to simultaneously model the interaction between contextual semantics and structural language features, thereby uniformly optimizing term boundaries at the sequence level.

To address the issue that the BERT tokenizer may split multi-word terms into subword units, this study employs a word-level alignment strategy for handcrafted features processing. Specifically, in the preprocessing stage, the original word sequence is first obtained by space segmentation, and then part-of-speech tagging and case judgment are performed using SpaCy. Subsequently, the features of each original word (such as POS tags and initial capitalization) are broadcast to all its corresponding BERT subword units. For example, if "technical textiles" is split into ["technic", "##al", "textiles"], then all three subwords inherit the original word-level features of "technical" and "textiles". Word frequency and positional features are also calculated and mapped at the original word granularity. This strategy effectively avoids the loss or misalignment of sub-word level features, ensuring that the structural information of multi-word terms is completely transmitted to the model.

D. DEFINITION OF LABEL SYSTEM AND FORMULATION OF TERM CLASSIFICATION STANDARDS

To achieve fine-grained recognition and structured annotation of English terms in higher vocational courses, this paper constructs a multi-dimensional and hierarchical label system, and combines the BIOES annotation specification to support nested term structure recognition. The system covers two levels: subject category and term type, forming an extensible and reusable term classification framework.

In terms of label design, the BIOES annotation format is used to finely model the term boundaries: let the input sequence be $W = (w_1, w_2, \ldots, w_T)$, where T is the sentence length, and the label $y_t \in \{B, I, O, E, S\}$ corresponding to each token w_t represents the start (Begin), inside (Inside), non-term (Outside), end (End), and single-word term (Single) of the term. This mechanism effectively solves the problem that the traditional Begin, Inside, Outside (BIO) format is difficult to handle nested structures. For any continuous term segment $[i, j]$, the label sequence satisfies $y_i = B$, $y_{i+1}, \ldots, y_{j-1} = I$, and $y_j = E$. For single-word terms, it is marked as $y_t = S$. On this basis, a multi-dimensional label space for term classification is constructed. The subject categories are divided into six categories: Computer Science, Electrical Engineering, Mechanical Engineering, Business Administration, Healthcare, and Automotive Technology, denoted as $C_{\text{subject}} = \{c_1^s, c_2^s, \ldots, c_K^s\}$, where $K = 6$, and combined with the term boundary labels to form composite labels such as B-Computer Science, I-Electrical Engineering, etc. According to the semantic role, the term type is divided into algorithm, device, process, and theory, denoted as $C_{\text{type}} = \{c_1^t, c_2^t, c_3^t, c_4^t\}$. For example, the term "decision tree" is classified as Algorithm, "voltage regulator" as Device, "supply chain management" as Process, "Ohm's Law" as Theory. By jointly encoding the term type label and the boundary label, a label structure such as S-Algorithm, B-Device is formed.

E. DESIGN OF MODEL OPTIMIZATION OBJECTIVE FUNCTION

To alleviate the problem of uneven distribution of category samples in the English terminology recognition task of higher vocational courses and improve the recognition rate of low-frequency terms, this paper introduces a weighted cross entropy loss function in the model training stage, and combines the label transfer penalty mechanism in the CRF decoding process to construct a multi-level optimization target.

Supposing that the term classification task contains K categories of labels, the number of samples of the i-th category term in the training set is n_i, and the total number of samples is $N = \sum_{i=1}^{K} n_i$. The category weight coefficient α_i is defined as follows:

$$\alpha_i = \frac{N}{K \cdot n_i} \tag{15}$$

In the BERT-based sequence classification task, the model outputs the predicted probability distribution $p_t = (p_{t1}, p_{t2}, \ldots, p_{tK})$ corresponding to each token, where p_{tk} represents the probability that the t-th token belongs to the k-th class. Let the true label be y_t, then the weighted cross entropy loss function is defined as follows:

$$L_{\text{WCE}} = -\frac{1}{T} \sum_{t=1}^{T} \alpha_{y_t} \log p_{t,y_t} \tag{16}$$

In addition, the label imbalance processing mechanism is further integrated in the CRF decoding layer. Let the label transfer matrix be $T \in R^{K \times K}$, where T_{ij} represents the score transferred from label i to label j. To suppress the suppression effect of high-frequency labels on low-frequency labels, the transfer score is dynamically adjusted during the training process:

$$T'_{ij} = \begin{cases} T_{ij} + \lambda \cdot \dfrac{\alpha_j}{\max(\alpha)}, & i \neq j, \\ T_{ij}, & i = j. \end{cases} \tag{17}$$

In formula 17, λ is the transfer gain coefficient, which is used to control the penalty intensity. This mechanism increases the transfer preference of low-frequency labels and guides the model to prefer low-frequency term labels when decoding, thereby alleviating the prediction bias caused by the difference in sample quantity.

F. CONSTRUCTION OF REAL-TIME TERM ANNOTATION ENGINE FRAMEWORK

To meet the real-time and engineering requirements of automatic annotation of English terms in the processing of higher vocational teaching resources, this paper designs and implements a lightweight term annotation engine system. In terms of system backend construction, Python language combined with FastAPI framework is used to build RESTful API service, and its asynchronous characteristics are used to improve the concurrent request processing capability. The interface receives the input text list in JavaScript Object Notation (JSON) format and returns the structured term annotation results, including fields such as term text, location information, tag type, and subject affiliation.

To improve the inference efficiency, the trained BERT-CRF model is exported to TorchScript format, and the computational overhead is reduced through static graph optimization. Assuming that the model input sequence is $X \in R^{B\times L_{\max}\times d}$, B is the batch size; $L_{\max} = \max_{b=1}^{B}(L_b)$ is the maximum length after dynamic padding; d is the feature dimension. At the same time, the attention mask matrix $M \in \{0,1\}^{B\times L_{\max}}$ is introduced to shield the influence of the padding position:

$$M_{b,t} = \begin{cases} 1, & t \le L_b, \\ 0, & t > L_b. \end{cases} \tag{18}$$

During the deployment process, a cache strategy is introduced; a consistent hashing algorithm is used to generate a unique text key value k = MD5(s); the result is stored in the Redis database. At the front-end interaction level, a Vue.js-based Web application is built to provide file upload, term highlighting, and classification statistics functions. Users can upload documents by dragging and dropping, and the system background calls the batch processing interface for parsing, and renders the term results with color markings on the page for intuitive visualization.

III. EXPERIMENT AND VERIFICATION

A. EXPERIMENTAL DESIGN

To verify the effectiveness of the English term automatic tagging engine (TermTag-BERTCRF) model based on BERT-CRF proposed in this paper in the scenario of higher vocational courses, a complete experimental evaluation system is constructed. First, this paper builds a higher vocational English term dataset based on multisource teaching resources (Table 1).

TABLE 1. Basic information of the dataset

Data Dimension	Description
Corpus Source	Covers various teaching resources such as textbooks, courseware, test papers, and lab manuals, collected from multiple vocational colleges and school-enterprise cooperative institutions.
Subject Coverage	Mainly includes six vocational education disciplines: Computer Science, Electronic Engineering, Mechanical Engineering, Business Administration, Healthcare, and Automotive Technology.
Text Type	Includes English textbook chapters, PPT lecture slides, past exam questions, training guides, and equipment specification documents.
Term Category	Terms are categorized into four types: Algorithm-related (Algorithm), Device-related (Device), Process-related (Process), and Theory-related (Theory)

As shown in Table 1, the data set comes from a variety of teaching resources such as textbooks, courseware, test questions, and experimental manuals provided by multiple higher vocational colleges and school-enterprise cooperation units. In terms of subject distribution, it focuses on the six key professional directions of higher vocational education: Computer Science, Electrical Engineering, Mechanical Engineering, Business Administration, Healthcare, and Automotive Technology, reflecting the professional breadth and industry relevance of data collection. The texts are in various forms, including English textbook chapters, PPT handouts, past exam questions, practical training guides, and equipment manuals, which can effectively reflect the characteristics of terminology usage in different teaching links; the terminology classification adopts a four-category system, namely algorithm, device, process and theory. To ensure the transparency and reproducibility of the assessment, this study constructs a dataset of higher vocational English terminology containing 12,847 annotated documents and a total of 2,153,694 tokens, of which 86,412 are manually annotated. The terminology is evenly distributed across six professional fields, with an average document length of 167.6 tokens and a maximum document length of 2,104 tokens. All terms are cross-validated by three field instructors, resulting in a Kappa coefficient of 0.89 for annotation consistency. This scale is sufficient to support deep model training while avoiding overfitting due to data sparsity. To comprehensively evaluate the performance of the TermTag-BERTCRF model proposed in this paper, three mainstream sequence tagging models are selected as baseline comparison models, including BiLSTM-CRF [32,33], BERT-BiLSTM-CRF [34], and RoBERTa-BiLSTM-CRF [35,36]. The specific structure and performance of each model are shown in Table 2.

TABLE 2. Structural performance characteristics of each model

Model Name	Structure Features
BiLSTM-CRF	Based on traditional word vector input, uses BiLSTM to extract contextual features, and optimizes label decoding using a CRF layer.
BERT-BiLSTM-CRF	Uses BERT as the encoder to extract contextual representations, introduces BiLSTM on top of BERT to enhance sequence modeling capabilities, and combines CRF for decoding.
RoBERTa-BiLSTM-CRF	Replaces BERT with the stronger RoBERTa pre-trained model, maintaining the same structure as BERT-BiLSTM-CRF.
TermTag-BERTCRF	Conducts secondary pre-training on vocational education-related corpora based on BERT, integrates multi-level handcrafted features, and introduces a weighted loss function to optimize low-frequency term recognition.

As shown in Table 2, BiLSTM-CRF uses bidirectional LSTM to extract context features based on traditional word vector input, and optimizes label decoding through CRF; BERT-BiLSTM-CRF uses BERT as the encoder, combines BiLSTM to enhance sequence modeling capabilities, and is decoded by CRF; RoBERTa-BiLSTM-CRF uses the same structure but uses a stronger RoBERTa pre-trained model. TermTag-BERTCRF performs secondary pre-training on vocational education corpus based on BERT, integrates multi-level manual features, and introduces a weighted loss function to improve the recognition of low-frequency terms.

B. EVALUATION OF OVERALL RECOGNITION PERFORMANCE OF THE MODEL

Figure 2 shows the overall performance of each model in the task of English term recognition in higher vocational courses. The evaluation metrics include accuracy, recall, and F1 Score.

As can be seen from Figure 2, the accuracy, recall, and F1 value of TermTag-BERTCRF reach 92.4%, 89.7%, and 91.0%, respectively, which are significantly better than other comparison models. The accuracy, recall, and F1 value of BiLSTM-CRF are 86.8%, 83.4%, and 85.1%, respectively; the three indicators of BERT-BiLSTM-CRF are 90.6%, 87.6%, and 89.1%, respectively. The accuracy, recall, and F1 value of RoBERTa-BiLSTM-CRF are 91.8%, 89.0%, and 90.4%, respectively. In summary, these results demonstrate the advantages of the TermTag-BERT- CRF model over the other annotation models for English term recognition in higher vocational education scenarios. In particular, the higher recall and F1 scores indicate that the model identifies target terms more accurately while reducing missed detections.

C. FINE-GRAINED TERM RECOGNITION CAPABILITIES

Figure 3 shows the differences in recognition performance of each model for English terms of different lengths. The horizontal axis represents the term length, and the vertical axis reflects the recognition accuracy of the model at that length.

From the data shown in Figure 3, the accuracy of English term recognition of the BiLSTM-CRF model at different lengths is 88.2% (1 word), 83.6% (2 words), 76.4% (3 words), 68.9% (4 words), 61.2% (5 words), and 54.7% (6 words), showing an overall downward trend with increasing length. The BERT-BiLSTM-CRF model has improved the English term recognition accuracy at the corresponding lengths, which are 90.5%, 86.7%, 79.8%, 72.1%, 65.3%, and 58.6%, respectively, showing the improvement effect based on the pre-trained language model. The RoBERTa-BiLSTM-CRF model further improves the performance, with recognition accuracy of 91.3%, 87.9%, 81.2%, 74.5%, 67.8%, and 60.2% at the six lengths, indicating that stronger language modeling capabilities are helpful for term recognition. The TermTag-BERTCRF model proposed in this paper outperforms other models at all lengths, with corresponding recognition accuracy rates of 94.5%, 89.7%, 85.4%, 78.7%, 70.4%, and 64.9%, respectively. Overall, with the increase of term complexity, the recognition accuracy of each model generally decreases, but the decrease of TermTag-BERTCRF is relatively small, showing better fine-grained recognition ability than other models, further verifying its effectiveness and adaptability in automatic annotation of English terms in higher vocational courses.

D. PERFORMANCE EVALUATION OF LOW-FREQUENCY ENGLISH TERM RECOGNITION

Figure 4 shows the performance evaluation of each model in the low-frequency English term recognition task. The detection indicators include mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The horizontal axis represents different model names, and the vertical axis corresponds to the specific values of each indicator.

The TermTag-BERTCRF model shows the best performance in all indicators, with an MAE of 0.09, RMSE of 0.14, and MAPE of 10.6% in Figure 4. The basic model BiLSTM-CRF performs relatively poorly in various indicators, with an MAE of 0.15, a RMSE of 0.21, and an MAPE of 16.2%. The BERT-BiLSTM-CRF and RoBERTa-BiLSTM-CRF models perform at a mid-range level in various evaluation indicators. The MAE of BERT-BiLSTM-CRF is 0.13; the RMSE is 0.19; the MAPE is 14.8%. In contrast, the MAE of RoBERTa-BiLSTM-CRF drops to 0.12; the RMSE is 0.18; the MAPE drops to 13.5%.

Judging from the overall data trend, the model performance shows a trend of continuous improvement with the gradual evolution of the structural design: from BiLSTM-CRF to BERT to RoBERTa, various evaluation indicators have improved to varying degrees, showing the obvious advantages of pre-trained language models in semantic modeling. On this basis, TermTag-BERTCRF introduces a task-oriented improvement strategy, which effectively enhances the ability to capture the contextual features of low-frequency terms, thereby significantly improving the recognition accuracy and further reducing the error level.

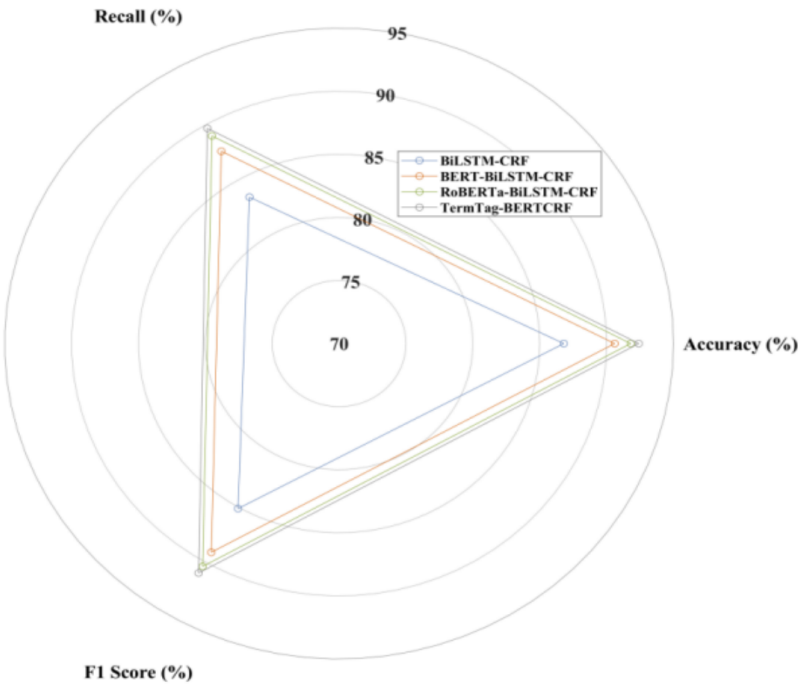

FIGURE 2. Overall recognition performance evaluation results of each model

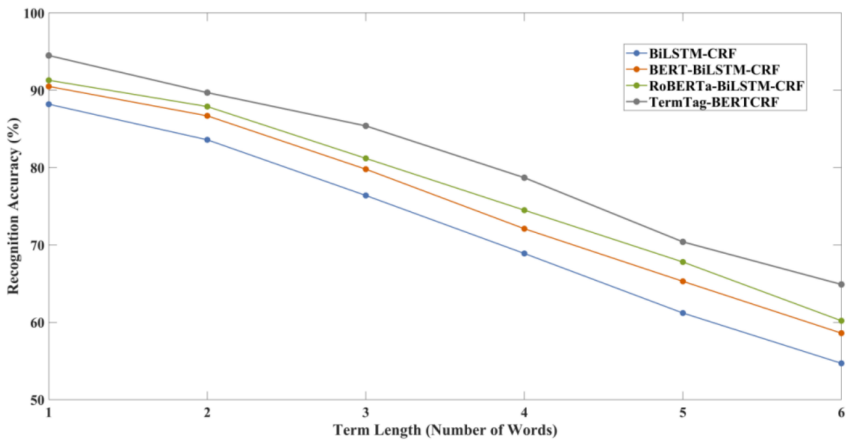

FIGURE 3. Analysis of fine-grained term recognition capabilities of each model

This result shows that combining BERT-like models with the CRF decoding mechanism, supplemented by task-driven structural optimization, is an effective means to address the challenge of low-frequency word recognition.

E. CONTEXTUAL DEPENDENCY SENSITIVITY

Figure 5 shows the analysis results of the contextual dependency sensitivity of each model, using Context Sensitivity Index (CSI) as the measurement indicator. The color depth reflects the CSI value. The higher the value, the stronger the model's dependence on and ability to utilize contextual information.

As shown in Figure 5, the TermTag-BERTCRF model has the best overall performance, with its CSI value gradually increasing from 0.86 to 0.92 and remaining high under most context lengths. The RoBERTa-BiLSTM-CRF model is second, with CSI values ranging from 0.84 to 0.88, showing strong context capture capabilities. The CSI value of the BERT-BiLSTM-CRF model is stable between 0.81 and 0.85, which is better than the BiLSTM- CRF model. The latter has the lowest CSI value, which only slowly increases from 0.76 to 0.80, showing the limitations of traditional neural networks in context modeling. The above results show that the TermTag-BERTCRF model can show higher stability and sensitivity when facing different context lengths, especially when dealing with common compound

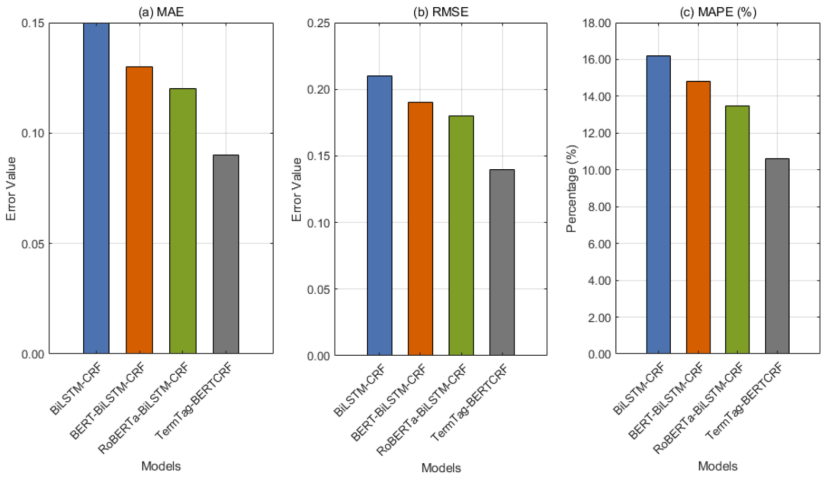

FIGURE 4. Analysis of the low-frequency English term recognition performance of each model; Figure 4a. MAE; Figure 4b. RMSE; Figure 4c. MAPE

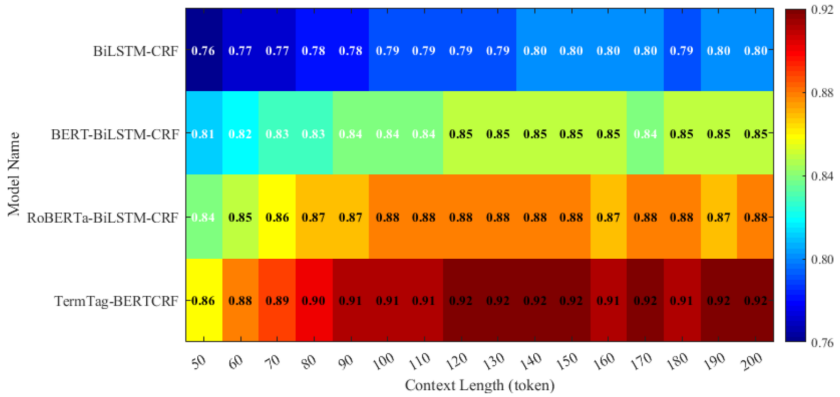

FIGURE 5. Analysis of context dependency sensitivity of each model

terms and nested structures in higher vocational courses. It can effectively use context information to improve annotation accuracy, reflecting its technical advantages in practical application scenarios.

F. ANNOTATION EFFICIENCY AND RESOURCE CONSUMPTION

Figure 6 compares the performance of each model in terms of annotation efficiency and resource consumption, using single document response time, batch throughput, first response delay, and peak memory usage as measurement indicators.

From the data in Figure 6, the TermTag-BERTCRF model is ahead of the comparison model in terms of processing efficiency and resource utilization. Specifically, its average single-document response time is 128.6 ms, which is much lower than BiLSTM-CRF (152.3 ms), BERT-BiLSTM-CRF (254.7 ms), and RoBERTa-BiLSTM- CRF (301.5 ms), showing higher operating efficiency. In terms of batch throughput, the model reaches 14.8 documents/second, significantly better than the other three groups of models, indicating that it is more adaptable under high-concurrency tasks. At the same time, its first response delay is 109.3 ms, which is also at the optimal level, indicating that the TermTag-BERTCRF model can better meet the application scenarios with high real-time requirements. In terms of resource usage, the peak memory of the TermTag-BERTCRF model is only 379.4 MB, which is much lower than BERT-BiLSTM-CRF (948.3 MB) and RoBERTa-BiLSTM-CRF (1098.7 MB), reflecting a good lightweight design. The above results show that the TermTag-BERTCRF model achieves higher annotation efficiency and lower resource overhead at the same time, which is very suitable for the automatic terminology annotation task in the scenario of higher vocational courses.

G. ROBUSTNESS AND NOISE TOLERANCE TEST

Figure 7 shows the changes in the F1 value of each model for higher vocational English term recognition under different noise ratio conditions, which is used to evaluate the robustness and stability of each model in a noisy environment. The horizontal axis represents the noise ratio (0%–25%), and the vertical axis represents the models. The color depth in the heat map intuitively reflects the F1 score of each model under different interference intensities.

From the data in Figure 7, under noise-free conditions, the

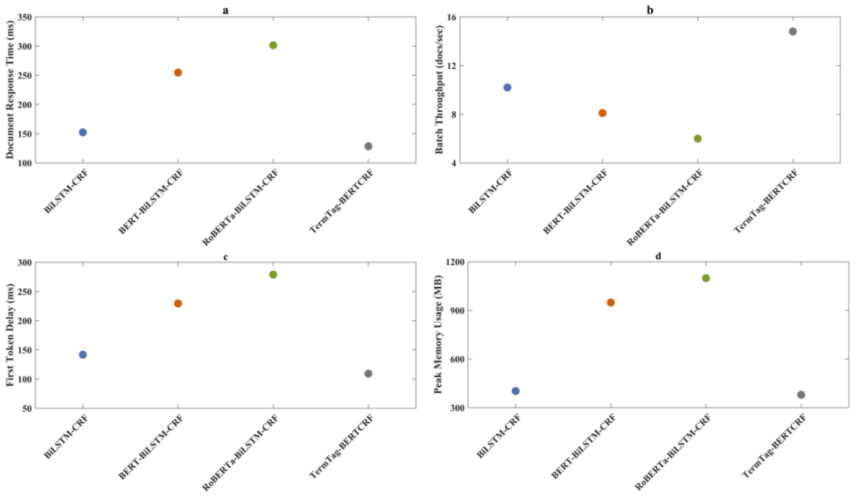

FIGURE 6. Annotation efficiency and resource consumption of each model; Figure 6a. Single document response time; Figure 6b. Batch throughput; Figure 6c. First response delay; Figure 6d. Peak memory usage

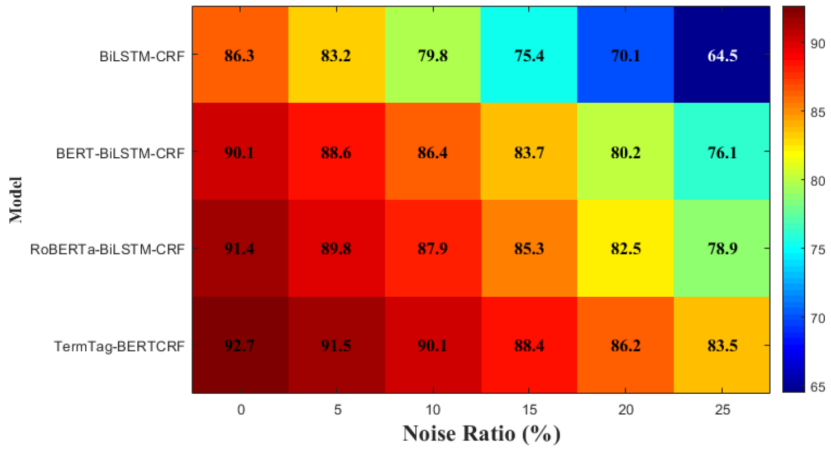

FIGURE 7. Robustness and noise tolerance test results of each model

F1 scores of the TermTag-BERTCRF, RoBERTa- BiLSTM-CRF, BERT-BiLSTM-CRF, and BiLSTM-CRF models are 92.7%, 91.4%, 90.1%, and 86.3%, respectively. As the noise ratio increases to 25%, the F1 scores of all models decrease to varying degrees. When the noise ratio reaches the highest 25%, the F1 score of the TermTag-BERTCRF model is 83.5%, which is still significantly higher than the 78.9% of the RoBERTa-BiLSTM-CRF model, the 76.1% of the BERT-BiLSTM-CRF model, and the 64.5% of the BiLSTM-CRF model. Overall, the performance degradation of the TermTag-BERTCRF model at various noise levels is relatively gentle, and it has stronger anti-interference capabilities. The performance of other models, especially the BiLSTM-CRF model, degrades more significantly in a high-noise environment, reflecting that its structure has a weak ability to model contextual information and is difficult to effectively deal with uncertainty in the input.

H. EVALUATION OF THE GENERALIZATION ABILITY OF CROSS-COURSE ENGLISH TERMINOLOGY RECOGNITION

Table 3 shows the evaluation results of each model in the generalization ability of cross-course English terminology recognition, covering six representative courses of higher vocational majors, including programming basics, circuit analysis, engineering mechanics, management principles, clinical nursing skills, and automotive electronics technology. The evaluation indicator is the F1 score.

TABLE 3. F1 Scores for Cross-Course Generalization of English Terminology Recognition

Course	BiLSTM-CRF	BERT-BiLSTM-CRF	RoBERTa-BiLSTM-CRF	TermTag-BERTCRF
Programming Fundamentals	76.3%	82.1%	84.5%	89.2%
Circuit Analysis	74.1%	80.3%	82.6%	88.0%
Engineering Mechanics	74.8%	80.6%	83.0%	88.5%
Principles of Management	70.5%	78.0%	81.1%	86.1%
Clinical Nursing Skills	72.1%	78.9%	81.2%	86.4%
Automotive Electronics	73.6%	79.7%	82.4%	87.6%

In Table 3, in the basic programming course, the F1 score of the TermTag-BERTCRF model reaches 89.2%, which is significantly higher than the 76.3% of the BiLSTM-CRF model, the 82.1% of the BERT-BiLSTM-CRF model, and the 84.5% of the RoBERTa-BiLSTM-CRF model. In the circuit analysis course, the F1 score of the TermTag-BERTCRF model is 88.0%, while the other three models are 74.1%, 80.3%, and 82.6%, respectively. In the engineering mechanics course, the F1 score of the TermTag-BERTCRF model is 88.5%, while the other models are 74.8%, 80.6%, and 83.0%, respectively. In the management principles course, the F1 score of the TermTag-BERTCRF model reaches 86.1%, compared with only 70.5% for the BiLSTM-CRF model, 78.0% for the BERT-BiLSTM-CRF model, and 81.1% for the RoBERTa-BiLSTM-CRF model. In the clinical nursing skills course, the F1 score of the TermTag-BERTCRF model is 86.4%, which is significantly better than the 72.1% of the BiLSTM-CRF model, the 78.9% of the BERT-BiLSTM-CRF model, and the 81.2% of the RoBERTa-BiLSTM-CRF model. In the automotive electronics technology course, the F1 score of the TermTag-BERTCRF model is 87.6%, which also exceeds the 73.6% of the BiLSTM-CRF model, the 79.7% of the BERT-BiLSTM-CRF model, and the 82.4% of the RoBERTa-BiLSTM-CRF model.

Overall, the performance of each model in different courses shows some differences, but the TermTag-BERT-CRF model is always at the leading level, and its advantages in multiple courses are particularly prominent, showing that the model has stronger adaptability and generalization ability when facing terminology in different professional fields.

IV. CONCLUSION

This paper proposes a TermTag-BERTCRF model based on BERT-CRF to meet the needs of automatic annotation of English terminology in higher vocational education courses. This model is particularly suitable for technical fields , accurately annotating professional terms in course materials and significantly im-proving the efficiency of education resource preparation. By pre-training BERT twice on a higher voca-tional education corpus, combining a BiLSTM-CRF decoding module and multi-level manual feature fusion, and introducing a weighted cross-entropy loss function and a label transfer penalty mechanism, the model's performance in fine-grained terminology recognition, low-frequency terminology recognition, and cross-course generalization is effectively improved. Experimental results show that the TermTag-BERTCRF model exhibits stable high performance in six core higher vocational courses (such as programming, circuit analysis, and clinical nursing), with F1 scores ranging from 86.1% to 89.2%. This indicates that the model has good cross-course generalization ability within the vocational education fields covered by the pre-training corpus. However, this generalization ability is limited by the professional scope of the secondary pre-training corpus, and its transfer effect in completely unseen fields (such as high-energy physics or clinical medical monographs) has not yet been verified. Therefore, the "interdisciplinary generalization" referred to in this paper should be strictly understood as the adaptability of multiple disciplines within vocational education, rather than zerosample transfer across general domains. Future work can explore domain-adaptive fine-tuning strategies to extend to broader engineering and technical subfields, such as manufacturing and design.

AUTHOR CONTRIBUTIONS

Conceptualization – Jingjing Zhu and Tao Song; methodology – Jingjing Zhu, Tao Song, Xiaoran Li and Liping Zhang; investigation – Jingjing Zhu and Tao Song; writing-original draft preparation – Jingjing Zhu, Tao Song, Xiaoran Li and Liping Zhang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was funded by Shandong "14th Five-Year" Education Science Planning Project (General Project): "Research on the Development and Construction of Digital Resources for English in Higher Vocational Education from the Perspective of "Vocational Education Going Global'" (No. 2023YB068)

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] R. Ye and E. Nylander, "Conventions of skilling: The plural and prospective worlds of higher vocational education," European Educational Research Journal, vol. 23, no. 5, pp. 709-727, 2024, doi: 10.1177/14749041241262826.

[2] M. Mei, "The contribution of higher vocational education to regional economic growth," International Journal of Innovation and Sustainable Development, vol. 19, no. 2, pp. 172-187, 2025, doi: 10.1504/IJISD.2025.144826.

[3] Z. Li, J. Zheng, and J. Xiong, "Examining project-based governance of higher vocational education in China: A case study," Higher Education Policy, vol. 36, no. 2, pp. 250-269, 2023, doi: 10.1057/s41307-021-00251-z.

[4] M. J. Tividad, "The English language needs of a technical vocational institution," Mary Joy Tividad (2024). The English Language Needs of a Technical Vocational Institution. Psychology and Education: A Multidisciplinary Journal, vol. 16, no. 3, pp. 256-275, 2024, doi: 10.2139/ssrn.4695007.

[5] N. K. Yıldırım and N. G. Bal, "English Language Education in a Vocational School: A Qualitative Case Study," Participatory Educational Research, vol. 10, no. 2, pp. 275-297, 2023, doi: 10.17275/per.23.40.10.2.

[6] L. Hao, K. Tian, U. K. Mohd Salleh, C. H. Leng, S. Ge, and X. Cheng, "The effect of project-based learning and projectbased flipped classroom on critical thinking and creativity for business English course at higher vocational colleges," Malaysian Journal of Learning and Instruction (MJLI), vol. 21, no. 1, pp. 159-190, 2024, [Online]. Available: https://repo.uum.edu.my/id/eprint/30356.

[7] M. G. V. N. V. Kumar, L. Čepová, M. A. M. Raja, A. Balaram, and M. Elangovan, "Evaluation of the quality of practical teaching of agricultural higher vocational courses based on BP neural network," Applied Sciences, vol. 13, no. 2, pp. 1180-1189, 2023, doi: 10.3390/app13021180.

[8] H. Jiang and Y. Liu, "Construction of teaching quality evaluation system of higher vocational project-based curriculum based on CIPP model," International Journal of Information and Education Technology, vol. 11, no. 6, pp. 262-268, 2021, doi: 10.18178/IJIET.2021.11.6.1521.
[9] J. Zhu and D. Wang, "Research on Curriculum Reform of Environmental Art and Design Majors in Higher Vocational Education Based on School Enterprise Cooperation," Advances in Vocational and Technical Education, vol. 5, no. 8, pp. 75-80, 2023, doi: 10.23977/avte.2023.050813.
[10] A. Picoral, S. Staples, and R. Reppen, "Automated annotation of learner English: An evaluation of software tools," International Journal of Learner Corpus Research, vol. 7, no. 1, pp. 17-52, 2021, doi: 10.1075/ijlcr.20003.pic.
[11] L. Lei and H. Wang, "A Multilabel Learning-based Automatic Annotation Method for Semantic Roles in English Text," IEEE Access, vol. 11, pp. 106220-106231, 2023, doi: 10.1109/ACCESS.2023.3319384.
[12] X. Fei, "An LDA based model for semantic annotation of Web English educational resources," Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 3445-3454, 2021, doi: 10.3233/JIFS-189382.
[13] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," Multimedia tools and applications, vol. 82, no. 3, pp. 3713-3744, 2023, doi: 10.1007/s11042-022-13428-4.
[14] M. A. Kazakova and A. P. Sultanova, "Analysis of natural language processing technology: Modern problems and approaches," Advanced Engineering Research (Rostov-on-Don), vol. 22, no. 2, pp. 169-176, 2022, doi: 10.23947/2687-1653-2022-22-2-169-176.
[15] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to deep learning in natural language processing: Models, techniques, and tools," Neurocomputing, vol. 470, pp. 443-456, 2022, doi: 10.1016/j.neucom.2021.05.103.
[16] S. Kaya, "From Needs Analysis to Development of a Vocational English Language Curriculum: A Practical Guide for Practitioners," Journal of Pedagogical Research, vol. 5, no. 1, pp. 154-171, 2021, doi: 10.33902/JPR.2021167471.
[17] N. E. Simbolon, "English Medium Instruction (EMI) in Higher Education: Insights from Indonesian Vocational Lecturers," Utamax: Journal of Ultimate Research and Trends in Education, vol. 5, no. 1, pp. 11-20, 2023, doi: 10.31849/utamax.v5i1.9973.
[18] Y. Li, "Construction of Internet of Things English terms model and analysis of language features via deep learning," The Journal of Supercomputing, vol. 78, no. 5, pp. 6296-6317, 2022, doi: 10.1007/s11227-021-04130-7.
[19] D. Wang, J. Su, and H. Yu, "Feature extraction and analysis of natural language processing for deep learning English language," IEEE Access, vol. 8, pp. 46335-46345, 2020, doi: 10.1109/ACCESS.2020.2974101.
[20] X. Guo, "Research on logical structure annotation in English streaming document based on deep learning," Journal of Computer Science, vol. 32, no. 4, pp. 109-122, 2021, doi: 10.53106/199115992021083204009.
[21] J. Li, Y. Shi, and S. Li, "Analysis of Beijing Traffic Violations Based on the BERT-CRF Model," Promet-Traffic&Transportation, vol. 36, no. 2, pp. 279-293, 2024, doi: 10.7307/ptt.v36i2.366.
[22] X. Wang, Y. Zhu, H. Zeng, Q. Cheng, X. Zhao, H. Xu, et al., "Spatialized analysis of air pollution complaints in Beijing using the BERT+ CRF model," Atmosphere, vol. 13, no. 7, pp. 1023-1035, 2022, doi: 10.3390/atmos13071023.
[23] A. S. Fitri, S. F. A. Wati, H. H. Putra, S. Widodo, and A. A. Aziiza, "Entity Extraction and Annotation for Job Title and Job Descriptions Using Bert-Based Model," Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, vol. 10, no. 1, pp. 73-77, 2025, doi: 10.25139/inform.v10i1.7367.
[24] K. Ma, M. Tian, Y. Tan, Q. Qiu, Z. Xie, and R. Huang, "Ontology-based BERT model for automated information extraction from geological hazard reports," Journal of Earth Science, vol. 34, no. 5, pp. 1390-1405, 2023, doi: 10.1007/s12583-022-1724-z.
[25] H. Chouikhi, M. Alsuhaibani, and F. Jarray, "BERT-based joint model for aspect term extraction and aspect polarity detection in Arabic text," Electronics, vol. 12, no. 3, pp. 515-524, 2023, doi: 10.3390/electronics12030515.
[26] Q. Zhang, C. Xue, X. Su, P. Zhou, X. Wang, and J. Zhang, "Named entity recognition for Chinese construction documents based on conditional random field," Frontiers of Engineering Management, vol. 10, no. 2, pp. 237-249, 2023, doi: 10.1007/s42524-021-0179-8.
[27] A. Pradhan and A. Yajnik, "Parts-of-speech tagging of Nepali texts with Bidirectional LSTM, Conditional Random Fields and HMM," Multimedia Tools and Applications, vol. 83, no. 4, pp. 9893-9909, 2024, doi: 10.1007/s11042-023-15679-1.
[28] S. Warjri, P. Pakray, S. A. Lyngdoh, and A. K. Maji, "Part-of-speech (POS) tagging using conditional random field (CRF) model for Khasi corpora," International Journal of Speech Technology, vol. 24, no. 4, pp. 853-864, 2021, doi: 10.1007/s10772-021-09860-w.
[29] R. Siddalingappa and K. Sekar, "Bi-directional long short term memory using recurrent neural network for biological entity recognition," IAES International Journal of Artificial Intelligence, vol. 11, no. 1, pp. 89-101. http://doi.org/10.11591/ijai.v11.i1.pp89-101, 2022.
[30] A. K. Nandanwar and J. Choudhary, "Semantic features with contextual knowledge-based web page categorization using the GloVe model and stacked BiLSTM," Symmetry, vol. 13, no. 10, pp. 1772-1788, 2021, doi: 10.3390/sym13101772.
[31] P. Baruah, B. Dutta, S. K. Sarma, and K. Talukdar, "Named Entity Recognition in Assamese Language using two separate models: BiLSTM and BERT," Procedia Computer Science, vol. 258, pp. 242-251, 2025, doi: 10.1016/j.procs.2025.04.262.
[32] F. Meng, S. Yang, J. Wang, L. Xia, and H. Liu, "Creating knowledge graph of electric power equipment faults based on BERT–BiLSTM–CRF model," Journal of Electrical Engineering & Technology, vol. 17, no. 4, pp. 2507-2516, 2022, doi: 10.1007/s42835-022-01032-3.
[33] B. Cai, S. Tian, L. Yu, J. Long, T. Zhou, and B. Wang, "ATBBC: Named entity recognition in emergency domains based on joint BERT-BILSTM-CRF adversarial training," Journal of Intelligent & Fuzzy Systems, vol. 46, no. 2, pp. 4063-4076, 2024, doi: 10.3233/JIFS-232385.
[34] D. Contractor, B. Patra, and P. Singla, "Constrained BERT BiLSTM CRF for understanding multi-sentence entityseeking questions," Natural Language Engineering, vol. 27, no. 1, pp. 65-87, 2021, doi: 10.1017/S1351324920000017.
[35] H. Zhang, Q. Du, Z. Chen, and C. Zhang, "A Chinese address parsing method using RoBERTa-BiLSTM-CRF," Geomatics and Information Science of Wuhan University, vol. 47, no. 5, pp. 665-672, 2022.
[36] B. Li, H. Cheng, and M. Lin, "SESG-Optimizing Information Extraction in Chinese Clinical Texts: An Innovative Named Entity Recognition Approach Using RoBERTa-BiLSTM-CRF Mechanism," Journal of Information & Knowledge Management, vol. 23, no. 06, pp. 2450090-2450111, 2024, doi: 10.1142/S0219649224500904.