International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1–18

https://doi.org/10.70917/ijcisim-2026-0224
Article

Impact of a Knowledge Graph Technology Supported English Terminology Learning System on Mastery of English in Disciplines

Lin Gong*

* Hulunbuir University, Hulunbuir, Inner Mongolia, 021008, China; ggl852123@126.com

Abstract: In the face of the challenge of fragmented knowledge in English, the study starts from helping students master terminology knowledge in a more systematic way, and constructs an English terminology learning system driven by knowledge graph technology. By combining top-down and bottom-up approaches, the study utilizes the pre-trained BERT-BiLSTM-CRF model to extract and construct a knowledge graph for the subject of English at university. The system integrates key technologies such as efficient resource indexing, online collaborative text-based cross-domain mapping extension, and time-ordered knowledge inference. Firstly, 32703 knowledge point entities are mined from 20 core courses, and the accuracy of entity alignment can reach up to 95.38%, while the average F1 value of recognizing various types of entities is also stable at about 94.40%. In actual teaching, the experimental class using this system constructed in this paper, the performance improvement is very obvious - the average score directly rose from 68 to 86, to a certain extent, narrowing the achievement gap between students, and there is no low-scoring students. At the same time students were more active in the classroom, with the level of assessment of learning increasing from 3.58 to 4.77, and the intention to learn behavior increasing dramatically from 2.87 to 4.56. Satisfaction with the system and instructional design even reached a high score of 4.78 and 4.64 respectively. Regression analysis shows that the system itself explains 8.1% to 14.5% of the improvement in each English proficiency.

Keywords: knowledge graph; English terminology; learning system; entity recognition; knowledge reasoning

1. Introduction

In today's increasingly deep globalization, the importance of English language teaching has become more and more prominent, but the current teaching is generally faced with the challenges of single content, traditional methods, and low student interest[1]. As mankind enters the age of intelligence, the deep integration of artificial intelligence and education promotes educational change and innovation, and realizes the development of high-quality education and teaching [2]. In 2012, Google formally put forward the concept of "knowledge graph", which was initially intended to improve the search engine [3]. Knowledge graph technology significantly optimizes the user search effect and experience by enhancing the understanding of the information retrieval system [4]. In essence, this technology builds a network structure with semantic association characteristics, and its core features are reflected in the two aspects of semantic parsing and structured presentation, while integrating the dual attributes of graphical display and systematic organization [5-6]. The use of knowledge graph technology is not only conducive to improving the intelligent level of the teaching mode, but also provides students with a personalized learning experience, thus enhancing their learning efficiency [7-8].

The construction of knowledge graph focuses on the acquisition of a large number of entities and entity relationships, in which the extraction of entity relationships is an important basis for data knowledge mapping, and the extraction of entity relationships belongs to the domain of information extraction [9]. Therefore, knowledge mapping has been widely used in major domains and different knowledge networks have been established. In the medical field, literature [10] established the world's largest neuroscience database at Yale University and mapped the neural network knowledge about brain structure. Literature [11] extracted medical concepts from a large number of anonymous patient records and built a knowledge graph in medicine using three probabilistic models such as logistic regression, which enabled the construction of a high-quality knowledge base associating diseases with symptoms. The new framework of “Safe Drug Recommendation” proposed in [12] constructs a high-quality heterogeneous graph through electronic medical records and medical knowledge graph, and embeds diseases, drugs, patients, and their corresponding relationships in a low-dimensional space to realize effective drug recommendation while avoiding drug-drug interactions. Literature [13] constructed a quaternion-structured medical knowledge graph through eight steps of data preparation, entity identification, entity normalization, relationship extraction, attribute computation, graph cleaning, related entity ranking and graph embedding, which combined with an entity ranking function can effectively learn the embedded representations of all the entities and relationships from large-scale e-cases. In the social domain, literature [14] extracts users' social relationships through graph convolutional neural networks, and then uses multi-task learning and semantic-based matching model to extract the information of attribute knowledge graph, while integrating the social relationships and commodity attribute data effectively improves users' retrieval efficiency. Another domain application of knowledge graph is that complex biological systems are modeled as interconnected biological entity graphs, i.e., biological knowledge graphs [15]. Literature [16] investigated the new knowledge embedding model for biological knowledge graphs and their different applications, and by introducing the knowledge embedding model, complex biological knowledge graphs can be processed and analyzed more efficiently, which brings more possibilities and inspirations for biological research and applications.

Educational knowledge graph is the application of knowledge graph in education field, whose nodes are knowledge concepts or knowledge points with educational significance, drawn according to the semantic relationship to form a knowledge network graph, which is mainly used for resource management, knowledge navigation and so on [17]. Literature [18] points out that knowledge graphs play an important role in enriching personalized learning, curriculum design, concept mapping, and educational content recommendation systems, and these structured knowledge representations are indispensable tools. Literature [19] attempts to allow learners to deepen their learning of knowledge points by constructing, modifying and refining, and querying and comparing knowledge graph nodes on their own during the teaching and learning process as a way to enhance students' knowledge mastery. Literature [20] constructed a comprehensive knowledge graph covering multidisciplinary knowledge points, and then combined with big data technology to analyze the user's historical behavior and learning preferences, so as to construct a dynamic learning path recommendation system, which significantly improved the user's course completion rate. Literature [21] proposes a method of applying knowledge mapping to online teaching, which helps learners plan personalized learning paths more accurately by guiding them to observe the connections between knowledge points, thus improving the overall learning efficiency. Literature [22] constructed a multimodal knowledge graph by collecting teachers' knowledge lectures in the classroom and incorporating speech data into the nodes of the knowledge graph, and experimentally proved the feasibility of the method used in managing and representing various types of educational resources.

Literature [23] takes the improvement of English teaching as the starting point, aims at the current problems in traditional English classroom teaching and online learning, constructs a knowledge graph with English-related knowledge points as the data source, creates user portraits based on the data of students' online learning behaviors on the learning platform, and builds a personalized English learning resource recommendation system based on the knowledge graph in order to satisfy their personalized needs and improve their English learning efficiency. Literature [24] constructed a specific knowledge graph for English learning based on the contextual information of the nodes, and the nodes on the knowledge graph are stable and coherent, which provides a systematic knowledge network structure for students, and the knowledge graph improves the existing teaching problems and enhances the quality of teaching. Literature [25] combed the fragmented online learning resources through the use of knowledge graph, combined with intelligent recommendation technology to provide students with personalized learning resources, the technology's resource recommendation relevance score of 4.41 points, and significantly improved students' English performance. Literature [26] combines the knowledge graph embedded scoring algorithm with the link scoring algorithm to design and build an optimized English knowledge graph quiz system for English teaching, and the application of this system to carry out teaching experimental research, which confirms that it can effectively solve the problem of missing questions in the existing knowledge graph embedded quiz method, and has good generalization performance. It can be seen that there are fewer research results on knowledge graph in the field of English teaching, and the few existing studies mainly focus on the development and construction of personalized resource recommendation system and Q&A system. Therefore, it is necessary to further explore its potential in the innovation of teaching mode and enhancement of teaching effect.

The study introduces the knowledge mapping technology, organizes the massive scattered English subject knowledge, and constructs an English terminology learning system that can help students master the knowledge in depth. Teaching resources are first digitally processed and organized, including the extraction of core terms and features. On the basis of setting up an efficient resource indexing system, an English knowledge map is established to connect abstract concepts and concrete entities through semantic relationships to form a semantic network. Further, incorporating online collaborative learning interaction data to supplement the mapping, entities are automatically identified and relationships are extracted from students' interaction texts, realizing the expansion of cross-domain knowledge mapping, so as to make it more comprehensively reflect the real learning paths and cognitive states. Ultimately, knowledge reasoning gives the system a certain intelligent inference ability, and at the same time introduces the temporal point process model to explore the implicit logic and temporal evolution law among the knowledge, to realize the completion of missing information and the prediction of future learning trends, and to further enhance the system's intelligence level.

2. A knowledge graph-based approach to constructing an English terminology learning system

2.1. Design of Knowledge Graph-based Digital Teaching Resources Retrieval Method for English Courses

2.1.1. Extracting the characteristics of digital teaching resources

According to the actual teaching needs, after the construction scope of the digital teaching resources ontology for college English courses is divided, the core terms in the digital teaching resources for college English courses are strictly screened and sorted out to extract the core terms, ensuring that each term can accurately reflect an important aspect of the course content. Then, the feature extraction design of digital teaching resources is carried out, and the process is shown in Figure 1.

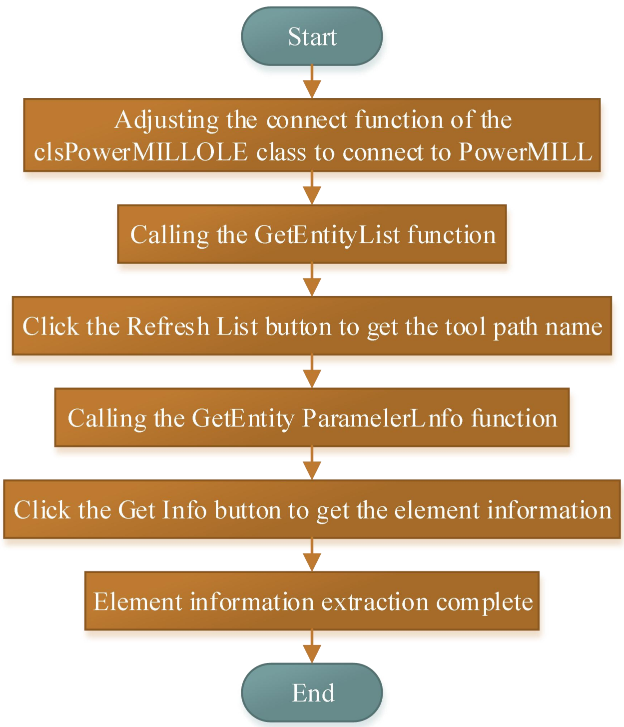

Figure 1. Process of extracting characteristics of teaching resources

On this basis, the extracted data features of teaching resources are processed to draw the spatial map of data feature structure. The assessment indexes of the same data features with homogeneity are known to quantify their space, and the formula is calculated as

$$A=\frac{\sum_{c=1}^{d}|h|^{c}\left(d-d_{c}\right)^{2}+\sum_{c=1}^{d}H^{c}\left(d+d_{c}\right)^{2}}{2b} \tag{1}$$

In Eq. (1), A represents the result of feature structure space quantization; h represents the difference assessment index; d represents the data feature item; H represents the homogeneity assessment index; b represents the retrieval cycle; and c represents the total number of features. After completing the spatial quantization of data feature structure, the quantization results are sampled by sparse coding and processed by dimensionality reduction to retain the representative features in order to reduce the redundant information and improve the efficiency of subsequent processing.

2.1.2. Indexing and feature data detection design

In order to realize efficient management and convenient retrieval of teaching resources for English courses in colleges and universities, so as to improve teaching quality and learning experience, a resource indexing system is constructed based on advanced information technology and database management technology. This time, a relational database management system is adopted and a flexible database schema is designed to support diversified information of resources (e.g., name, size, time cutoff, file format, etc.).

In addition, NLP-based word segmentation technology slices text resources into word sequences and removes duplicate words through hash tables and other techniques to reduce the index storage space occupied. Backward index entries are created for each different word, containing information about the word itself, the resource ID and its location in the document to support fast full-text search. The teaching resource indexing relationship formula is

$$F=\frac{X_{c}(A)\times Y(A)}{\beta} \tag{2}$$

In Eq. (2): F represents the information index item; X represents the feature data ID locator; Y represents the uniform resource locator; and β represents the semantic interval. Based on the above, the establishment of teaching resources index for college English courses is completed. Next, the deep learning algorithm is used to obtain the semantic similarity rate of literature search, so as to analyze the fuzzy boundaries between the complex semantic concepts involved in the feature data and the direct meanings of the keywords, and then test the relevance of the index to the feature data after removing the highly similar data. The sub-vectors of the feature data are analyzed, and the feature data are regarded as the core component of the similarity analysis, and the selected features are combined into a feature vector N, which represents the position of the data points in the feature space. Based on this result, the similarity is calculated for different feature data with the formula

$$\sin\alpha=\frac{N_{1}\times N_{2}}{\left|N_{1}\right|\times\left|N_{2}\right|}=\frac{\sum^{F}N_{1F}\times N_{2F}}{\sqrt{\sum^{F}\left(N_{1F}\right)^{2}}\times\sqrt{\sum^{F}\left(N_{2F}\right)^{2}}} \tag{3}$$

In Eq. (3): α represents the vector pinch angle of the feature data. The similar data in the pinch angle are organized, one of them is selected and saved, and all the rest are deleted. Using point-to-point information technology, it is organized into the amount of information carried by the feature data and the index, so as to assess the relevance, which is calculated as follows

$$PIM\left(X,F'\right)=\log\frac{u\left(X,F'\right)}{u(X)\times u\left(F'\right)} \tag{4}$$

In Eq. (4): $PIM\left(X,F'\right)$ represents the point mutual information correlation value; F' represents the collapsed feature information; and u represents the correlation probability obtained by maximum likelihood estimation. It is known that $u\left(X,F'\right)/u(X)\times u\left(F'\right)=1$ when the feature information is independent of the index; the higher the correlation between the two $PIM\left(X,F'\right)$, the larger it is. According to this result, the feature data with unique properties are corresponded to the index to realize the detection of index and feature data correlation.

2.2. English Knowledge Mapping Construction in Higher Education

The previous section focuses on the digitization of English course resources and describes how to extract key features from massive teaching materials and establish an efficient indexing mechanism. In order to integrate these resources into a semantically related knowledge network, the construction of a knowledge graph for English in colleges and universities will be carried out next.

The construction of a knowledge graph consists of two main approaches: “top-down” and “bottom-up”. “Top-down” refers to the use of structured databases to define the ontology knowledge layer, and then gradually refine the entities to be added to the concepts. “Bottom-up” refers to the use of open link data to summarize and organize the entities, and then gradually form the upper layer of concepts. The University English Knowledge Atlas adopts a combination of “top-down” and “bottom-up” methods in the construction process.

The ontology knowledge layer is to extract and summarize concepts and the interrelationships among them from textbooks and teaching materials, forming a "concept - relationship - concept". Taking the "Professional Terminology Learning Course" in the college English knowledge graph as an example, the manual construction method is adopted. The knowledge points are regarded as concepts, and the knowledge points and their hierarchical relationships are extracted. Since knowledge points belong to the same category of concepts, the hierarchical relationship between them is a subcategory relationship. The chapter names and main text titles of teaching materials can, to a certain extent, reflect the knowledge points of the course. For instance, in the constructed knowledge graph of college English professional terms, terms such as "Science Field" and "Quantum mechanics field" mainly originate from chapter names, and the latter belongs to the subclass relationship of the former. The edges marked with "rdfs: subclass of" are used to present the subclass relationship between knowledge points.

Under the constraints of the ontology knowledge layer, the data resource layer, based on the specific knowledge points in the teaching materials, constructs a hierarchical structure including "entity - attribute - attribute value" and "entity - relationship - entity" by extracting the "attribute-value" relationship, entities and their mutual relationships. On this basis, by linking concepts with entities and integrating the ontology knowledge layer with the data resource layer, a college English knowledge graph is constructed. Continuing to take the English "Professional Terminology Learning Tutorial" as an example, for the knowledge points of professional terms in the medical field, the manual construction method is adopted. From textbooks, teaching syllabuses, courseware and other materials, "Cardiac Surgery" is extracted as an example, which is used as an entity in the data resource layer, and the edge marked with "rdf: type" is used to link the knowledge points and examples. In addition, the attributes and their attribute values of "Cardiac Surgery" are extracted. The entities and attribute values are linked through the edges marked by the attributes, demonstrating the semantic relationship between the concept and the entity, and forming a representation of "entity - attribute - attribute value".

In order to extend the coverage of knowledge elements and the application potential of the University English Knowledge Graph, raw data including unstructured data (e.g., pictures, audio, video, etc.) are incorporated into the construction process of the Knowledge Graph to extract the knowledge elements (e.g., knowledge points and re-difficulty levels) and to establish the associations between the knowledge elements and the resource objects. The micro-lesson videos utilized to determine the start time of a chapter's appearance and include the start time and URL as its attributes in the data resource layer to enhance the efficiency of knowledge resource acquisition. In addition, there may be instances of sharing between different knowledge points. For example, the human-computer combination method is used to search for the relevant content of a certain knowledge point in the textbook, courseware, syllabus and other materials of the “Tutorial on Learning of Technical Terms”, to construct a net-like knowledge structure between different instances and different knowledge points, and to form a fusion of knowledge, which effectively solves the problem of fragmentation of information in the process of acquiring and applying knowledge, and thus promotes the intelligence and efficiency of language learning.

2.3. Cross-domain Knowledge Graph Construction Techniques Based on Online Collaborative Learning Interactive Texts

The constructed knowledge graph of English terminology can help us realize the semantic links between concepts and entities. In order to integrate learners' language use in real contexts, the article continues to explore how to recognize entities and relations from the interactive text of online collaborative learning and realize the dynamic expansion of cross-domain knowledge graph.

The proposed technical route for constructing cross-domain knowledge graphs based on online collaborative learning interactive texts is shown in Figure 2, including entity recognition technology, relationship extraction technology and domain knowledge graph generation technology.

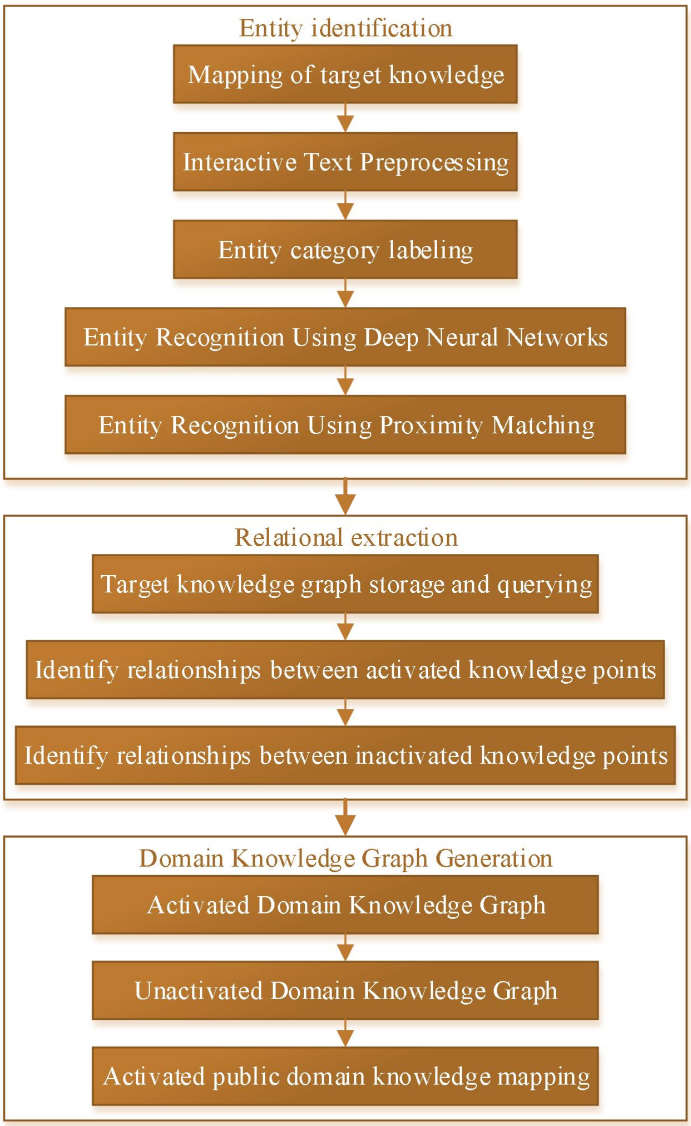

Figure 2. Construction of knowledge graph based on the interactive text

2.3.1. Entity recognition techniques

Combining the knowledge points in the target knowledge map and the online collaborative learning interaction text, this study divides the entity recognition technique into five steps.

(1) Determine and draw the target knowledge map according to the learning objectives and learning contents of different collaborative learning tasks. The target knowledge map consists of accurate and comprehensive target knowledge points and their relationships, covering all possible knowledge points and relationships during online collaborative learning interactions, as well as characterizing the complete knowledge structure and expected learning goals.

(2) Text pre-processing of the training data, including the removal of stop words, unnecessary symbols and expressions. The subject matter of the training data needs to be consistent with the subject matter of the actual online collaborative learning task to be carried out.

(3) Annotate the text with entity types. For the annotation of entity types, this study distinguishes three major types of entities, i.e., entities that need to be matched through semantic matching, entities that need to be matched using proximate words, and irrelevant topics, where entities that need to be matched through semantic matching can be further subdivided into different categories depending on the online collaborative learning task and the content of the interactive text.

(4) Recognizing entities that need to be matched by semantic matching using deep neural network models. In this study, three deep neural network models are mainly used for entity recognition, i.e.,

BERT, BERT-ISTM-CRF, and BERT-BiLSTM-CRF, and the model with the highest accuracy and prediction rate is selected as the final model for entity recognition. According to the previous exploration, BERT-BiLSTM-CRF has the highest accuracy rate, mainly because this model can recognize semantic contextual relationships.

(5) Entity recognition using near-synonym matching. It should be noted that the proximate lexicon needs to be created in advance for subsequent calls. Different proximal thesauri need to be created according to different collaborative learning tasks, and the proximal thesaurus should be accurate and complete.

2.3.2. Relational extraction techniques

Relationships in this study include two main types of relationships, i.e., relationships between activated knowledge points and relationships between inactivated knowledge points. Relationship extraction is mainly queried based on the relationships in the target knowledge graph, because the correct knowledge points and relationships are stored in the target knowledge graph. Relationship extraction is divided into three core steps.

(1) Store the knowledge points and their relationships in the target knowledge graph for relationship query.

(2) Identify the relationships between already activated knowledge points. This step first queries the activated knowledge points and then queries the relationships between the activated knowledge points from the target knowledge graph.

(3) Identify relationships between inactivated knowledge points. This step first queries the inactivated knowledge points, and then seeks the difference set based on the relationship between the target knowledge graph and the activated knowledge points, so as to identify the relationship between the inactivated knowledge points.

2.4. Intellectual reasoning

In order to read out more hidden information from the knowledge graph, the perspective now turns to knowledge reasoning, exploring what should be done to fill in and predict the missing content in the knowledge graph through temporal modeling and logical inference, so as to enhance the system's ability to comprehend and express the subject knowledge of English.

2.4.1. Knowledge Reasoning Task Definition

Knowledge graphs are often incomplete, in which case knowledge reasoning becomes particularly important. The purpose of knowledge reasoning is to infer new entities or relationships based on facts already in the knowledge graph through rule mining. Unlike traditional knowledge graphs, temporal knowledge graph reasoning (TKGR) can be categorized into reasoning within the observable time domain and reasoning outside the observable time domain, i.e., the two main categories of prediction and complementation, due to the addition of information in the time dimension.

In this context, prediction is the prediction of future facts, using interpolative reasoning, relying on the unidirectional character of time, and predicting possible future events or relationships by learning the evolutionary laws of historical knowledge. Given a temporal knowledge graph $G=(E,R,T)$ and historically known facts $\left\{\left(s,r,o,t_i\right) \mid t_i<t_q\right\}$, the goal is to predict, based on the historical facts, the missing entities in the incomplete quaternion on the future temporal truncation t_q $\left(s,r,?,t_q\right)$ or the relation $\left(s,?,o,t_q\right)$ for inference.

Completion, on the other hand, works on reasoning and supplementing facts in the observable time domain to improve the data and information in the graph to enhance the performance of downstream applications. It adopts an interpolative inference model to infer the missing parts by analyzing the historical and future information of the knowledge to be complemented. Given a temporal knowledge graph $G=(E,R,T)$, the task requires to complete the missing quaternions in the temporal knowledge graph, i.e., to the missing entities in the quaternions $\left(s,r,?,t_q\right)$, the relation $\left(s,?,\mathrm{o},t_q\right)$, and the time intercept $(s,r,o,?)$ for inference, where $t_q \in T$.

2.4.2. Timing point process

A temporal point process is a statistical tool for modeling and analyzing sequences of events. In reasoning about temporal knowledge graphs, such methods model the occurrence of facts as temporal point processes and calculate the probability of a particular event occurring in the future based on this historical information.

Know-Evolve models the occurrence of facts as a time-ordered point process by constructing a conditional intensity function to compute the probability of the next event. The conditional strength function is adjusted according to the fact's relationship score, which is determined by the dynamic embedding of the entity. Specifically, for a given set of historical events $T=\{t_1,\ldots,t_n\}$, the survival function $S(t\,|\,T)$ denotes the conditional probability that no event occurs within the time interval $[tn,t)$. Given a set of historical events $T(t)$, the conditional intensity function $\lambda(t)$ describes the probability of a new event occurring within the time interval $[t,t+dt)$. The probability of an event occurring at moment t is computed by means of the conditional intensity function $\lambda(t)$ and the survival function $S(t)$, i.e., $f(t)=\lambda(t)S(t)$. The conditioning of the conditional strength function relies on the outer product of the entity embedding vectors and the relation weight matrix R_r , which is computed as shown in Figure 3. These embedding vectors are dynamically updated to reflect the state of the entity prior to the current time. By combining the temporal point process and relational representation learning, Know-Evolve effectively captures the temporal dependencies between facts. By dynamically updating the entity embeddings and relationship weights, the model is able to model the evolutionary relationships between entities and thus predict the time of occurrence of facts.

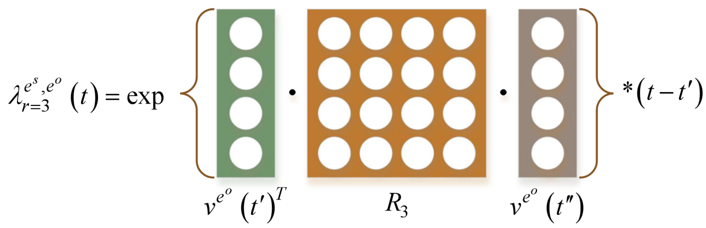

Figure 3. Intensity computation at time t

3. Validation of knowledge graph construction effect and entity recognition performance

After completing the construction of the English terminology learning system, the next step is to verify the actual effect of the system in two parts. The first is to assess the quality of the knowledge graph construction, examining whether its knowledge points are comprehensively extracted and whether the correlation is in place; the second is to quantify the performance of the entity recognition technology based on the online collaborative learning of interactive text.

3.1. Entity extraction and alignment analysis based on multi-source course resources

3.1.1. Course Resource Entity Extraction

With reference to the professional cultivation program of English discipline in universities, 20 English-related courses are selected for the construction of curriculum resource knowledge graph of university English discipline, and the electronic reference textbooks and lecture PPTs of these courses are mainly selected as source data. The pre-trained BERT-BiLSTM-CRF model was utilized to extract knowledge point entities from the constructed English curriculum resources knowledge graph, and a total of 32703 knowledge point entities were obtained. The specifics of the number of knowledge point entities for each curriculum resource in the English subject knowledge graph are shown in Table 1 below.

Table 1. The number of knowledge point entities for each course resource

Course Resources	Number of knowledge points	Course Resources	Number of knowledge points
English Professional Terms	2633	Selected Readings in English Newspapers	508
English Listening Comprehension	646	Analysis of English Films	268
English Oral Expression	1394	English Speeches and Debates	1123
English Reading Comprehension	2136	Societies and Cultures of Western Countries English Rhetoric	1402
English Writing Skills	1917	English Cross-Cultural Communication	2056
English Translation Theory and Practice	2429	English Lexicology	1576
Detailed Explanation of English Grammar	3073	Selected Readings in British and American Literature	3884
Introduction to English Linguistics	1403	Research on Western Drama	1004
Business English	2104	English Poetry Appreciation	714
Academic English Writing	1881	Selected Readings in English Newspapers	552

The number of knowledge entities contained in each course varies greatly, with “English Vocabulary” containing the largest number of knowledge entities, 3884, followed by “English Grammar” (3073) and “English Terminology” (2633), indicating that English terminology courses contain a large number of knowledge points in the knowledge system, reflecting the fact that only by mastering the knowledge related to English terminology can we continue to study in the English subject. Terminology" (2,633), indicating that English terminology courses contain the majority of knowledge points in the knowledge system, reflecting the fact that only by mastering the relevant knowledge of English terminology can we continue to work in the study of English.

3.1.2. Aligned Entity Detection

Based on the English Subject Knowledge Mapping data obtained from Table 1, the discussion is now divided into two groups of categories for the entity alignment of curriculum resources between PPTs and textbooks and the entity alignment between cross-curriculum categories, respectively. The course categories are divided into I: the English terminology category containing lexicography, academic English writing, etc.; II: the English language core skills category with listening comprehension, oral expression, etc. as the main courses; III: the literature and culture category and IV: the English application category emphasizing the use of English in the practical field. The specifics of the dataset of college English major courses and their number of aligned entities are shown in Figure 4 below.

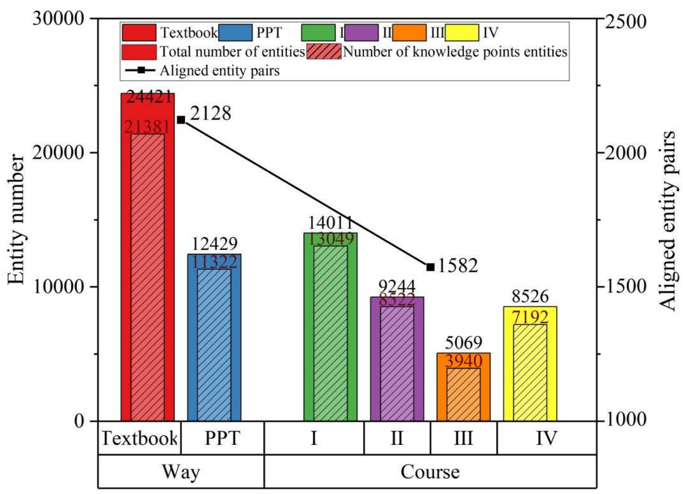

Figure 4. The number of aligned entities of English major course dataset

From the perspective of teaching aids, the number of knowledge entities contained in the resource map of teaching materials is 21,381, which is significantly more than the 11,322 in the resource map of PPT, indicating that the knowledge coverage of teaching materials is still more systematic and complete. From the perspective of course categories, "English terminology" occupies the largest number of knowledge entities with 13,049, which again proves the core position of terminology learning in the whole English knowledge system. The number of aligned entities detected by the model in the experiments on the dimensions of teaching aids and course categories are 2128 and 1582 respectively, which shows that the application of the aligned entity technology based on indexing and feature data is still effective.

3.1.3. Comparison Experiments of Alignment Algorithms

In order to further test the performance of this paper's method of aligning entities based on index and feature data, seven entity alignment algorithms based on knowledge representation learning are introduced for comparison.

(1) Evaluation Metrics

There are three mainstream ways to evaluate the performance of entity alignment algorithms, which are Hit@n series, MR and MRR.

Hit@n is the size of the proportion of correctly aligned entities appearing in the top n positions of the result ranking, if the value of Hit@n is larger, it means that the entity alignment method is more effective. In general, the values of m are set to 1, 3, 10 and 50.

Mean Rank MR: Reflects the average of the median result ranking of the correctly aligned entities, i.e., how many of them are averaged to match the correct result in the test set. If the MR value is smaller, it means that the entity alignment method is more effective.

Mean Reverse Rank MRR: Reflects the average value of the median reverse of the median result ranking of the correctly aligned entities, if the MRR value is larger, it indicates that the entity alignment method is more effective.

In this paper, we choose four evaluation indexes, Hit@1, Hit@10, MR and MRR, to evaluate the alignment entity method.

(2) Comparison models

The following seven entity alignment models are selected for comparison experiments.

MTransE: Use the translation model TransE to map two knowledge graphs into two spaces respectively, and then use the transfer matrix to transform the vectors in the two spaces into the same space.

BootEA: The same translation model TransE is used to embed the two knowledge atlases into the same space, and the accumulation of errors in the iterations is reduced by alignment editing.

JAPE: uses the translation model TransE and Skip-gram model to obtain structural embeddings and attribute embeddings, respectively, while capturing the correlation between attributes to compute the similarity matrix.

AttrGNN: Divide each knowledge graph into four subgraphs, jointly encode attribute and structural information, use graph neural network to obtain structural embedding, and assign different weights to different attributes through the attention mechanism.

HGCN: Combines graph convolutional network and high-speed gating mechanism, adopts high-speed gated graph convolutional network to learn the embedding representation of entities, and acquires the relationship representation vector based on head entity embedding and tail entity embedding.

RNM: Learning by mutual iteration of two entity alignment tasks and relationship alignment tasks, utilizing semantic information and mapping properties of relationships to uncover aligned entities.

OntoEA: exploits the hierarchical structure information that ontologies naturally have for entity alignment, using a graph structure representation that models the hierarchical structure in the mapping as a nonlinear transformation between a subclass and its associated parent class

Now 40% of the aligned entity pairs of the dataset are input as seed entity pairs, and the remaining 60% of the dataset is used as the test set for the experiment. The results of the performance of this paper's model and the above entity alignment model on the English course dataset are shown in Table 2.

Table 2. The performance results on the English major course dataset

	Textbook and PPT resource				**Different course categories resource**			
	Hit@1	Hit@10	MR	MRR	Hit@1	Hit@10	MR	MRR
MTransE	30.66	50.72	252	0.368	28.41	44.03	283	0.411
BootEA	38.53	55.57	310	0.324	43.67	50.54	317	0.346
JAPE	44.68	57.24	155	0.463	35.52	67.21	188	0.579
AttrGNN	64.81	70.51	51	0.631	68.27	76.19	47	0.641
HGCN	71.41	76.55	68	0.537	72.12	80.49	44	0.693
RNM	75.62	86.09	18	0.807	77.68	79.74	22	0.851
OntoEA	70.16	89.61	22	0.737	79.73	82.95	45	0.755
OURS	88.76	95.38	7	0.917	85.23	96.59	11	0.877

From the above comparative data, it can be seen that the method proposed in this paper demonstrates obvious advantages in all key indicators. In the alignment tasks of textbook and PPT resource graphs, the method proposed in this paper achieved 88.76% of Hit@1 values and 95.38% of Hit@10 values. This means that when looking for correctly aligned entities, the correct answer can be found in the first result in nearly 90% of cases, and in more than 95% of cases, it can be found in the first ten results. In contrast, the Hit@1 value of traditional methods such as MTransE is only 30.66%, showing a significant gap. The average ranking MR Of the method proposed in this paper on the two datasets is 7 and 11 respectively, which means that in all detection tasks, this model matches the correct results on average at the 7th and 11th, with a very high position. However, the subperforming OntoEA can only match correctly after 22 search results in the textbook and PPT resource graph.

To more clearly demonstrate the performance differences of different algorithms in the entity alignment task, bubble plots of the relationship between Hit@10 and MRR were also drawn as shown in Figures 5 and 6. The vertical axis represents Hit@10, and the size of the bubble indicates the matching score MRR. The larger the bubble and the higher the data points, the better the ability to align entities under this model.

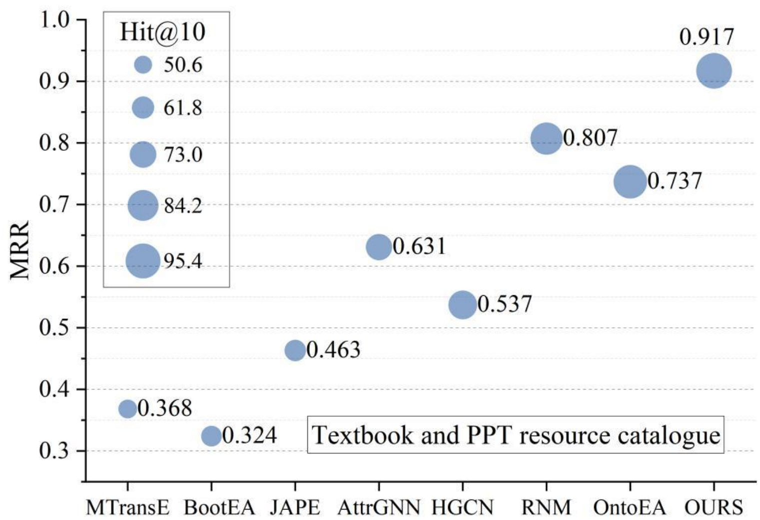

Figure 5. The Hit@10 and MRR bubble charts of the textbook and PPT resource

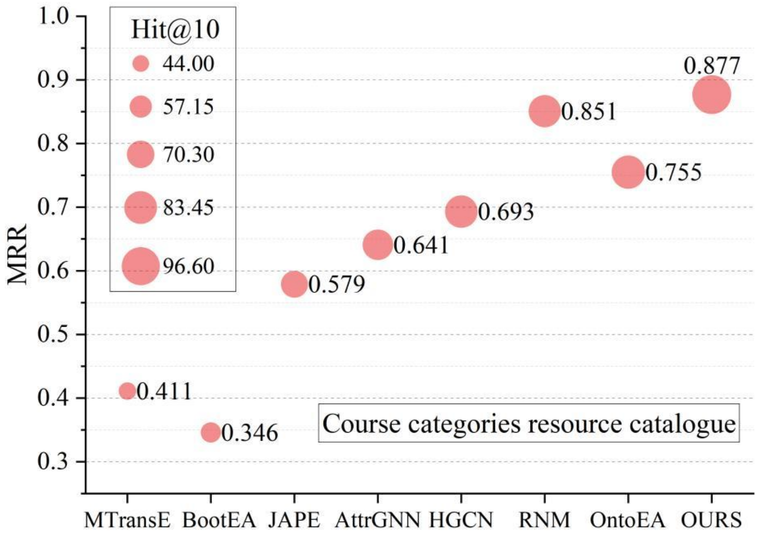

Figure 6. The Hit@10 and MRR bubble charts of different course categories

From the figure, it can be clearly seen that the bubble representing this paper's method not only has the highest position, but also has the largest bubble volume, i.e., both Hit@10 and MRR values are optimal, and this high and large feature graphically illustrates that this paper's method is able to align entities efficiently in indexing detection based on the features of digital teaching resources. In contrast, the bubble distribution of other algorithms is more dispersed, such as MTransE and BootEA, both of which have smaller bubbles and are concentrated in the lower-left region, indicating that their performances are relatively weak; while the position of RNM is high, but its bubble is not as large as that of OntoEA.

3.2. Cross-domain entity recognition model training and entity recognition effect analysis

After verifying the effectiveness of entity extraction and alignment in the knowledge graph construction process, we turn our attention to the performance results of the entity recognition model itself. Take a deeper look at how the core model, BERT-BiLSTM-CRF, performs in training and how accurate it really is in recognizing various types of entities.

3.2.1. Loss function

To further examine the performance of cross-domain entity recognition technique based on online collaborative learning of interactive text, the BERT-BILSTM-CRF model is now subjected to 100 iterations, and the loss function and model accuracy are shown in Fig. 7.

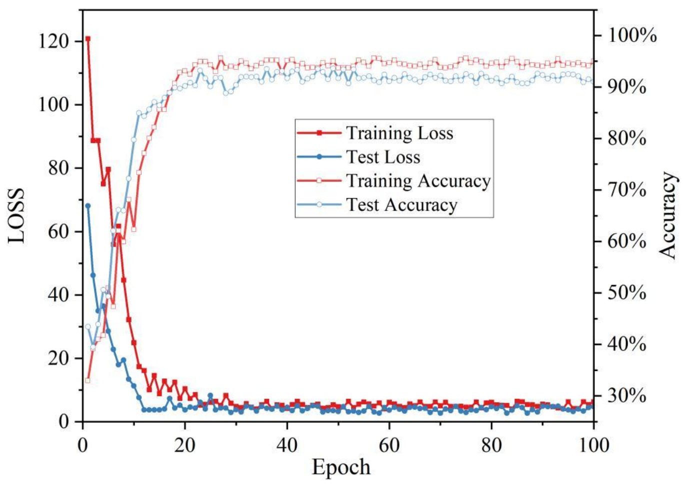

Figure 7. Loss function and model accuracy after 100 iterations

It can be seen that the loss value decreases with the number of iterations and reaches stability at almost 20 iterations. At the beginning of training, because the model parameters are randomly initialized, the model fits the data poorly at the initial stage, which can cause the loss value to be unstable. Meanwhile, the model accuracy curve shows a continuous upward trend and also reaches stability at almost 20 times, and the accuracy on the training set and test set stabilizes at about 94.37% and 91.68%, respectively, which indicates that the model is improving and is moving towards better prediction ability.

3.2.2. Comparison of different categories of entity recognition

The study adopts Accuracy, Precision, Recall and F1 value as evaluation metrics for the model's entity recognition ability. For MOOC data with high domain specificity, these four evaluation metrics are able to comprehensively analyze the model's recognition performance in that particular task. Figure 8 shows the entity recognition metrics of the BERT-BILSTM-CRF model under different categories.

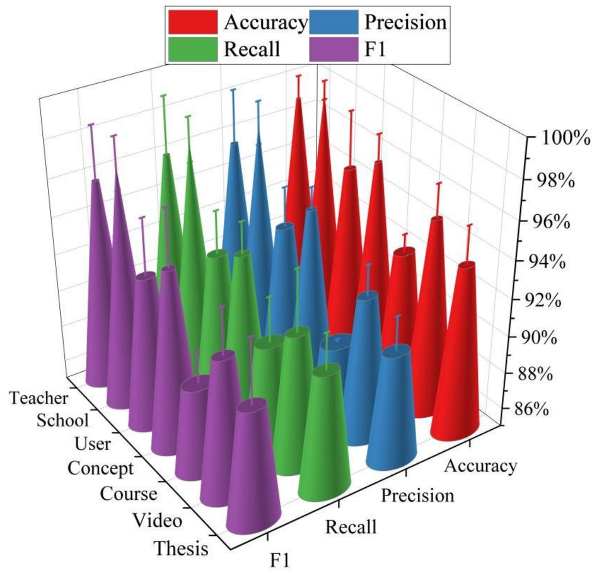

Figure 8. Recognition results of different categories based on BERT-BILSTM-CRF

The BERT-BiLSTM-CRF model performs well on different entity categories. The highest recognition accuracy of 98.21% was achieved for the category of “school”, which indicates that the model is capable of recognizing relatively fixed entities such as school names. In contrast, the “course” category is slightly less well recognized, with an accuracy of 92.49% and an F1 value of 89.62, probably because course names tend to be more variable, and recognition techniques based on the interactive text of online collaborative schools are not able to capture these variations well. Alternatively, the method in this paper has some deficiencies in understanding the interactive text information of the “course” entity, and is unable to accurately grasp the characteristics of the “course” entity, thus affecting the accuracy of the recognition. In contrast, the interactive textual information of the “concept” and “video” entities provides more clues and improves the recognition. Overall, the model's F1 value on all seven entity categories exceeds 89%, with an average of 94.40%. This performance is still quite stable, indicating that the BERT-BiLSTM-CRF model is indeed better adapted to the task of recognizing different categories of entities.

4. The Practice of Teaching English Terminology Based on Knowledge Graph Learning System

Let's shift our perspective from technology validation to teaching practice, and see what this knowledge mapping learning system can achieve in a real English classroom through a semester-long teaching experiment. We will examine the practical benefits of this new model around the core dimensions of student achievement, classroom performance, learning attitudes and satisfaction.

4.1. Design of teaching experiments

In this study, a quasi-experimental research method was used to select two classes of English majors in the class of 2024 in a university in city A to carry out teaching experiments, class A is the experimental class, which adopts the design mode of teaching English terminology based on the Knowledge Graph Learning System, and class B is the control class, which is taught using the traditional teaching mode. There are 52 students in both classes. The data of the experiment were collected through questionnaires, examination papers, classroom observation and interviews with teachers and students, etc. Based on the results of the statistical analysis of the experimental data, the effectiveness of the teaching mode of English terminology based on the knowledge map learning system was verified from the four aspects of the students' classroom performance, their learning attitudes, the degree of improvement of their academic performance, and the degree of their satisfaction with the teaching design and learning system.

4.2. Analysis of the results of the teaching experiment

4.2.1. Analysis of learning achievements

The pre-test and post-test in terms of English terminology were administered to the students of the two classes before and after the experiment was carried out, and Figure 9 shows the statistical results of the students' academic performance before and after the experiment in the two classes.

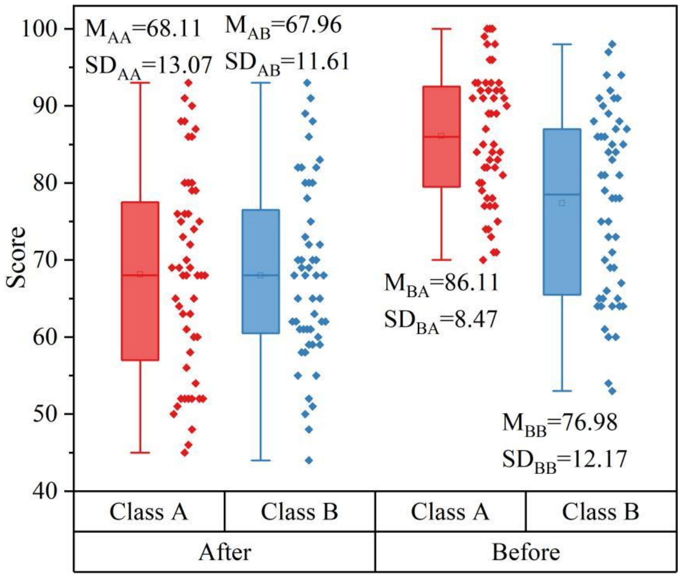

Figure 9. The academic performance of class A and B before and after the experiment

Before the start of the experiment, the average scores of the experimental and control classes stood on the same starting line, 68.12 and one 67.96, respectively, with a difference of less than 0.2 points, and the statistical test also showed that there was really no significant difference between the two groups. After one semester of instructional intervention, the average score of the experimental group using the Knowledge Graph Learning System increased to 86.10, which is nearly 18 points higher than before the experiment. The distribution of specific student grades in Figure 9 shows that there was a significant increase in the number of students scoring in the 90-100 range, and the lowest score was only 68, whereas a significant number of students scored below 60 failing grades prior to the experiment, indicating that the system was very effective in improving the overall student population, especially for the lower grades. The standard deviation of the scores decreased from 13.10 to 8.47, which indicates that the achievement gap between students is narrowing. In contrast, although the control group, which follows the traditional teaching mode, has also made progress, reaching 76.98 points, the improvement is obviously smaller, indicating that the traditional teaching mode is far less effective than the knowledge map-based learning system in improving the students' performance in English terminology.

4.2.2. Analysis of Classroom Performance, Learning Attitude and Satisfaction

A five-point Likert scale (1=Strongly Disagree, 5=Strongly Agree) was used to quantitatively assess the students' performance in conjunction with classroom observation records and student questionnaires. Figure 10 shows the classroom performance, learning attitudes, and satisfaction performance with the instructional design and learning system of the experimental group after a semester-long model of teaching English terminology based on the knowledge graph learning system.

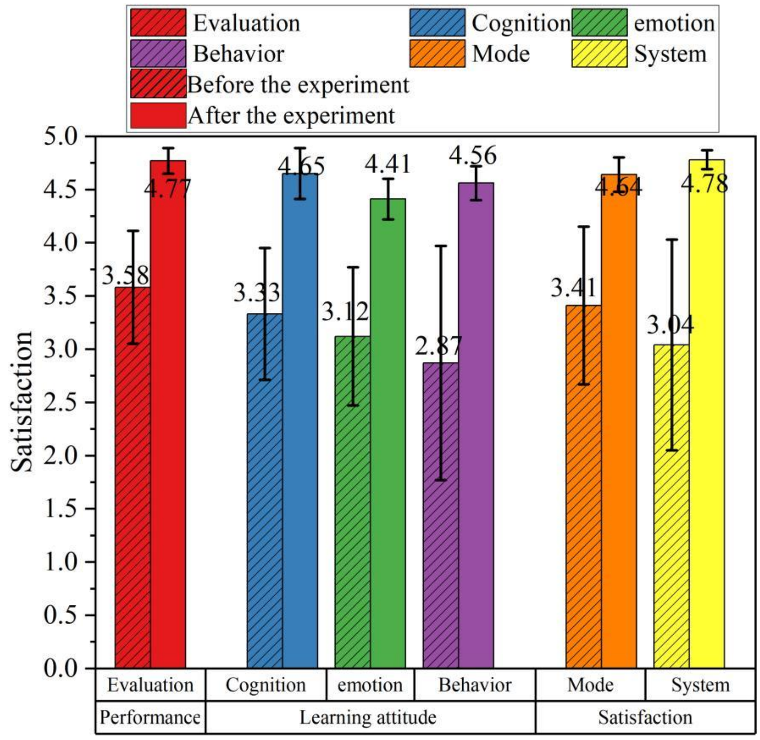

Figure 10. Analysis of Classroom Performance, Learning Attitude and Satisfaction

In terms of classroom performance, under the English terminology teaching mode, the students' learning assessment level jumped from 3.58 points before the experiment started to 4.77 points, a very solid increase. This shows that students are in a completely different state in the classroom, no longer passively listening to lectures, but more willing to actively participate in the exploration and discussion based on mapping, and the whole classroom's energy has been raised.

Looking at the learning attitudes, the changes in the three dimensions are also different. The biggest increase is in “learning behavior intention”, from the initial reluctance of 2.87 points to 4.56 points. This means that the system has successfully turned “I want to learn” into “I want to learn”, and students' willingness to use the system to learn has greatly increased. At the same time, their “Learning Perception” of English learning has also become much deeper, with the score increasing from 3.33 to 4.65, and their “Learning Affect” has also shifted from a little bit of resistance (3.12) to a little bit of interest (4.41).

Students' satisfaction with the learning system increased from 3.04 at the beginning to 4.78 at the end of the experiment, which is the highest score of all indicators. It can be said that students gave the new system a big thumbs up. Their satisfaction with the instructional design was also very high, at 4.64, indicating that they not only liked the tool itself, but also recognized the new teaching methodology behind it.

The system not only improves students' performance, but more importantly, it changes students' learning mindset and classroom ecology, so that they learn more actively, understand better and are more satisfied.

5. Regression Analysis of the Influence of English Mastery Proficiency of English Learners in Colleges and Universities

In order to explore to what extent exactly these positive changes can be directly attributed to the students' use of the system, we will dissect the data more finely with the help of multiple linear regression analysis. In this study, multiple linear regression analysis will be used to examine the role of the English terminology learning system supported by knowledge graph technology in influencing the mastery ability of English learners in higher education subjects. Since there are significant differences in some of the main variables among students with different genders, majors and places of origin, these three demographic variables are placed in the first block of the regression model as control variables in the regression analysis. Each variable of the English terminology learning system (terminology mastery, knowledge network integration, frequency of system use, and depth of system interaction) was placed in the second block of the model as the dependent variable. And then the 8 key competence variables of listening ability, reading ability, speaking ability, writing ability, discourse ability, learning ability, thinking ability, and cultural ability in English learning were sequentially put as dependent variables, and multiple linear regression analyses were done by forced entry method. The regression analysis of English terminology learning system on learners' mastery ability is shown in Table 3.

Table 3. Regression analysis of professional terms on the mastery ability

	R^2						β			
	Total	**Demographic variables**	**English major term learning system**	**D-W**	**F**	**df**	**Proficiency in professional terms**	**Integration degree of knowledge network**	**Frequency of system usage**	**Depth of system interaction**
Listening ability	.415/.400	0.028	0.105	1.972	27.845***	10,393	.217***	.184**	.095*	.158**
Reading ability	.388/.372	0.031	0.098	2.089	24.917***	10,393	.241***	.162**	0.073	.134*
Speaking ability	.451/.436	0.025	0.128	2.181	32.116***	10,393	.198***	.172**	.112*	.265*
Writing ability	.423/.408	0.029	0.116	2.045	28.774***	10,393	.283*	.201***	0.087	.154**
Pragmatic ability	.367/.350	0.033	0.092	2.031	22.653***	10,393	.176**	.155**	.101*	.118*
Learning ability	.528/.515	0.021	0.145	2.205	43.992***	10,393	.224***	.291*	.136**	.228***
Thinking ability	.335/.317	0.036	0.081	1.988	19.761***	10,393	.152**	.138*	0.064	0.097
Cultural ability	.509/.495	0.024	0.138	2.167	40.328***	10,393	.206***	.262***	.118*	.194***

The diagnostic results of multicollinearity show that the tolerance is between 0.472 and 0.948, both of which are greater than 0.2, and the maximum value of VIF is 2.024, which is less than 5, indicating that the problem of covariance in this regression model is negligible. The autocorrelation test results of the residuals show that the Durbin-Watson values are between 1.788 and 2.181, indicating that the serial correlation problem of the sample data basically does not exist. The results of ANOVA showed that the regression equation was valid with F (11823) = 17.018-39.372 ($p < 0.001$). After controlling for the effects of background factors such as gender and major, the module English Terminology Learning System alone explains an additional 8.1% to 14.5% of the variation in each competency. The most prominent contributions were made to the improvement of learning ability and cultural competence, with R2 values of 0.145 and 0.138 for changes in both.

Specifically, terminology mastery was key, and this metric had a significant positive effect on all eight competencies, especially with the largest effect on writing (β = .283) and reading (β = .241). This is understandable; the terminology itself is the building block of academic literacy, and the better a student's mastery of it, the more pronounced the improvement in performance will naturally be.

The dependent variable Knowledge Network Integration also performed impressively, with Beta values of .291 and .262 on improving academic and cultural competence, respectively.This is because knowledge mapping can help students connect fragmented knowledge points into a network, and this structured understanding can very much help students improve their cognitive and cultural insights.

Students' usage behavior is equally effective, and the depth of system interaction has a broader impact than mere frequency of use. It promotes speaking and learning skills in particular, with standardized regression coefficients Bata of 0.265 and 0.229 for both, suggesting that active inquiry-based learning behaviors are more likely to hone students' thinking and expression skills.

This English terminology learning system supported by knowledge graph technology constructed in the article is not just a simple vocabulary query tool, it improves students' comprehensive English literacy in all aspects of the subject by promoting their in-depth mastery of terminology and helping them to build a clear knowledge network in their minds.

6. Conclusion

The study accurately extracted more than 32,000 knowledge point entities from the massive course resources and constructed a knowledge graph of high quality. In the entity alignment experiment, Hit@10 reached a maximum of 95.38%, and the average F1 value in the recognition task was 94.40%.

All these prove that this technical solution is stable and reliable. Over the course of a semester, the grades of the experimental class using the knowledge graph-based English terminology system teaching model jumped by nearly 18 points, with an average score of 86.10, and the students' willingness to learn and their liking for the system increased dramatically, with the intention to learn behavior rising from 2.87 to 4.56. Regression analysis tells us that the system was able to independently explain, on average, 14.5% of the change in learning competence and cultural competence 13.8% increase. The article builds not only a vocabulary query tool, but also a learning partner that can guide students to build their personal knowledge system in deep understanding and knowledge association.

References

1. Du, L., Zhao, L., Xu, T., Wang, Y., Zu, W., Huang, X., ... & Wang, L. (2022). Blended learning vs traditional teaching: The potential of a novel teaching strategy in nursing education-a systematic review and meta-analysis. Nurse education in practice, 63, 103354.
2. Tedre, M., Toivonen, T., Kahila, J., Vartiainen, H., Valtonen, T., Jormanainen, I., & Pears, A. (2021). Teaching machine learning in K–12 classroom: Pedagogical and technological trajectories for artificial intelligence education. IEEE access, 9, 110558-110572.
3. Juel Vang, K. (2013). Ethics of Google's Knowledge Graph: some considerations. Journal of Information, Communication and Ethics in Society, 11(4), 245-260.
4. Li, J., Wang, Z., Wang, Y., Hua, Z., & Jing, W. (2021). Research on distributed search technology of multiple data sources intelligent information based on knowledge graph. Journal of Signal Processing Systems, 93(2), 239-248.
5. Malyshev, S., Krötzsch, M., González, L., Gonsior, J., & Bielefeldt, A. (2018, September). Getting the most out of Wikidata: Semantic technology usage in Wikipedia’s knowledge graph. In International Semantic Web Conference (pp. 376-394). Cham: Springer International Publishing.
6. Zhu, G., Hao, M., Zheng, C., & Wang, L. (2022). Design of knowledge graph retrieval system for legal and regulatory framework of multilevel latent semantic indexing. Computational Intelligence and Neuroscience, 2022(1), 6781043.
7. Dang, F., Tang, J., & Li, S. (2019, July). MOOC-KG: A MOOC knowledge graph for cross-platform online learning resources. In 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC) (pp. 1-8). IEEE.
8. Zhao, S. (2024, November). Personalized Learning Path Recommendation Strategy Based on Subject Knowledge Graph. In Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education (pp. 398-402).
9. Kong, F., Zhu, Q., & Zhou, G. (2012). Anaphoricity determination for coreference resolution in English and Chinese languages. Journal of Computer Research and Development, 49(5), 1072.
10. Grams, D. (2018). A quantitative study of the use of DreamBox learning and its effectiveness in improving math achievement of elementary students with math difficulties (Doctoral dissertation, Northcentral University).
11. Rotmensch, M., Halpern, Y., Tlimat, A., Horng, S., & Sontag, D. (2017). Learning a health knowledge graph from electronic medical records. Scientific reports, 7(1), 5994.
12. Gong, F., Wang, M., Wang, H., Wang, S., & Liu, M. (2021). SMR: medical knowledge graph embedding for safe medicine recommendation. Big Data Research, 23, 100174.
13. Li, L., Wang, P., Yan, J., Wang, Y., Li, S., Jiang, J., ... & Liu, Y. (2020). Real-world data medical knowledge graph: construction and applications. Artificial intelligence in medicine, 103, 101817.
14. Yang, G. A. O., & Yuan, L. I. U. (2023). Recommendation Algorithm Combining Social Relationship and Knowledge Graph. Journal of Frontiers of Computer Science & Technology, 17(1).
15. Alshahrani, M., Khan, M. A., Maddouri, O., Kinjo, A. R., Queralt-Rosinach, N., & Hoehndorf, R. (2017). Neuro-symbolic representation learning on biological knowledge graphs. Bioinformatics, 33(17), 2723-2730.
16. Mohamed, S. K., Nounu, A., & Nováček, V. (2021). Biological applications of knowledge graph embedding models. Briefings in bioinformatics, 22(2), 1679-1693.
17. Li, H. (2022). Piano Teaching Knowledge Graph Construction Based on Cross-Media Data Analysis and Semantic Network. Computational Intelligence and Neuroscience, 2022(1), 5499593.
18. Abu-Salih, B., & Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. Heliyon, 10(3).
19. Hanewald, R. (2012). Cultivating life-long learning skills in undergraduate students through the collaborative creation of digital knowledge maps. Procedia-Social and Behavioral Sciences, 69, 847-853.
20. Zhang, X. (2025). Research on the Application of Big Data Learning Recommendation Model Driven by Knowledge Graph Algorithm in Educational Information Platform Optimization. Systems and Soft Computing, 200298.
21. Xiao, G. (2023). A personalized learning path for French study in colleges based on a big data knowledge map. Scientific Programming, 2023(1), 4359133.
22. Li, N., Shen, Q., Song, R., Chi, Y., & Xu, H. (2022). MEduKG: a deep-learning-based approach for multi-modal educational knowledge graph construction. Information, 13(2), 91.
23. Huang, Y., & Zhu, J. (2021). A personalized English learning material recommendation system based on

knowledge graph. International Journal of Emerging Technologies in Learning (Online), 16(11), 160.
24. Wu, Z., & Jia, F. (2022). Construction and application of a major-specific knowledge graph based on big data in education. International Journal of Emerging Technologies in Learning (iJET), 17(7), 64-79.
25. Zhang, M., & Qiao, X. (2024, November). Combination of Knowledge Graphs to Make Personalized Recommendations for English Teaching Content. In Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education (pp. 368-373).
26. Wang, L. (2022). An improved knowledge graph question answering system for English teaching. Mobile information systems, 2022(1), 3401074.