International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 5452-5462

https://doi.org/10.70517/ijhsa464384

Application of Knowledge Graph Technology in Intelligent Management of Ideological and Political Education Content

Liming Tian[1,*]
[1] Marxist Theory and Ideological and Political Education, Central South University, Pingdingshan, Henan, 467000, China
Corresponding authors: (e-mail: tlm20252025@163.com).

Abstract The traditional way of organizing the content of Civic and Political Education relies on manual arrangement, which is inefficient and difficult to realize the analysis of inter-knowledge correlation. This paper proposes an intelligent organization and management method of Civic and Political Education content based on knowledge graph. The methodology adopts the Bert-Graph Attention-CRF model for entity recognition, and the BiGRU model that combines the attention mechanism with syntactic analysis to realize the relationship extraction, and constructs the Civic and Political Education Content Knowledge Graph. The study establishes a complete knowledge extraction process through semi-structured data crawling and unstructured text processing, and implements knowledge storage and visualization based on Neo4j graph database. The experimental results show that the F1 value of the proposed model reaches 96.23% in the Civics objective existence entity recognition task, and the F1 value is 88.46% in the Civics logical concept entity recognition. The system successfully acquires 5246 entities in the field of Civics and Politics and retains 746 valid entities after manual de-emphasis and merging, covering multiple categories such as political figures, political activities, political organizations and locations. The knowledge graph constructed in this study effectively solves the problem of intelligent organization and management of Civic and political education content, provides technical support for the informatization of Civic and political education, and significantly improves the utilization efficiency of teaching resources and knowledge discovery ability.

Index Terms Knowledge Graph, Civic and Political Education, Entity Recognition, Relationship Extraction, Attention Mechanism, Intelligent Organization

I. Introduction

Ideological and political education content is a basic element of the ideological and political education system, is the concretization of the ideological and political education purposes and tasks of the ideological and political education content structure is stable, scientific and reasonable, will be directly related to the effectiveness of ideological and political education [1], [2]. Ideological and political education content is a complex system, the elements of the system in accordance with a certain order, a certain way to combine into an organic whole, the stability and rationality of this whole is to ensure the effective implementation of the content of ideological and political education is an important condition, so it is important to carry out a reasonable organization and management to improve the effect of ideological and political education.

The traditional organization and management of teaching content is basically carried out by the subject teachers themselves, not only low efficiency but also increase the teaching pressure on teachers, while the ideological and political education is a course that advances with the times, and its course content will increase with the changes in society, in order to achieve the completeness and accuracy, artificial organization and management is obviously impractical [3]-[6]. And with the development of artificial intelligence, the application of knowledge graph brings intelligence to the organization and management of the content of ideological education [7], [8]. Knowledge graph is a data structure based on semantic network, which forms a complete knowledge system by associating and integrating the relationship between multiple knowledge points [9], [10]. In Civic Education, knowledge mapping is based on semanticization, structuring and other forms of organization, which realizes the organization of educational content and management efficiency, and can help students better understand knowledge, improve learning efficiency and teaching quality [11], [12].

This paper proposes a deep learning-based content knowledge extraction method for civic and political education, designs a Bert-Graph Attention-CRF entity recognition model and a relationship extraction model with attention mechanism fused with syntactic analysis, and constructs a professional knowledge graph in the field of civic and political education. The study first establishes a Civic and Political Education content corpus and adopts a multi-source data fusion strategy to obtain high-quality training data. Then an improved entity recognition algorithm is designed to effectively recognize conceptual entities and relationship types specific to the Civics and Political Science domain. Then we develop a relationship extraction model to accurately extract the semantic associations between entities. Finally, a knowledge map is constructed based on the extracted knowledge triples to realize the structured representation and visualization of the contents of Civic and Political education, forming a complete intelligent organization and management system.

II. Knowledge Extraction for Knowledge Mapping of Civic Education Content

In order to realize the intelligent organization and management of the content of ideology and politics education, this paper constructs the knowledge map of the content of ideology and politics education. This chapter analyzes the knowledge extraction of Civic and political education content and proposes the entity acquisition method and entity relationship extraction method of Civic and political content.

II. A. Methods of Acquiring Civic Content Entities

The practical application value of the constructed Civic Education Content Knowledge Graph is based on the knowledge extraction results, therefore, this paper proposes the entity naming recognition method of the Bert-Graph Attention-CRF model for the subsequent Civic Education Content Entity Extraction work.

II. A. 1) Feature embedding representation

The architecture of the Bert model consists mainly of a multilayered bi-directional Transformer structure, the basic principle of which is to obtain the feature vector representation of a word by evaluating its relevance to another word in the sentence, and then adjusting the matrix of weight coefficients to obtain a feature vector representation of the word.

(1) The text utterances are split into tokens sequences, and [CLS] and [SEP] symbols are added at the beginning and end of each text sequence respectively, which are processed into the input format of the model.

(2) The sequence input representation X_i consists of a vector of words themselves, a vector of features encoding positional information, and a vector of features used to differentiate between different sentences are summed, i.e:

$$X_i = tok_i + pos_i + seg_i \tag{1}$$

(3) After N Transformer layers, the feature distributed representation of X can be learned Y, formally mapped as:

$$f^{Bert} : X \to Y(X, Y \in R^{n \times d}) \tag{2}$$

where n denotes the number of characters in the Token and d denotes the dimension of the character embedding.

II. A. 2) Graph Composition

Graph composition is to model the extracted key information in a graph structure for the task requirements, this study uses the relationship between Chinese characters to represent the text in the form of topological graph structure, Chinese named entity recognition in this regard is converted into a node classification task.

After Bert's pre-training model, each character can form an embedding representation, which can be used to logically determine whether there is a relationship between each character and then convert the whole sentence into a directed graph form $G = (V, E)$, where $C_i \in V, e_i \in E$.

(1) Let a sentence be denoted as $S = \{C_1, C_2, C_3 \cdots C_n\}$, and C_i denotes the i th character feature in the sentence.

(2) Then the set of potential words that can be associated with the vocabulary in the sentence can be represented as $W_{b,e} = \{C_b, C_{b+1}, \cdots C_f\}$, where b is a shorthand of beginning, f is a shorthand of finish, b denotes the first character, f denotes the last character.

(3) If there is a potential word that can be matched, an edge $e_{b,f}$ on the potential word will be constructed.

II. A. 3) Attention Aggregation and Optimization

Graph structures make extensive use of the attention mechanism in the aggregation process, which aggregates graph information by assigning different weights to neighboring nodes or related edges. The attention layer allows the construction of features for each word, and unlike the recursive neural network form can assign different weights to each character vector of the input and incorporate the global semantics of the sentence into the vector representation of each word, being able to capture key information in the sequential input sequence.

The process of learning using the attention mechanism is as follows:

Step one. Define three linear transformation matrices Q(Query), K(Key), V(Values) to be learned, assuming that the dimensions of the three weight matrices W^Q, W^K, W^V and each character vector of the input X are all 512 dimensions, and the character vectors will be linearly computed to obtain the three matrices Q, K, V, and the dimension are all L*512 dimensions, where L is the number of characters in the sentence, and the computation process is shown in Equation (3):

$$Q = X * W^Q K = X * W^K V = X * W^V \tag{3}$$

Step two. The computation of the attention weights of the current word with respect to each word in the sentence is realized by the similarity computation between Query and each K, the similarity is denoted by f, i.e.:

$$f(Q, K_i) = QK_i^T \tag{4}$$

Step Three. Softmax normalization of the weights computed in Eq. (4) ensures that the weights of a row sum to 1. The Attention weights are then weighted and summed with the corresponding features, i.e., each word's weight is used to weight and sum each one-dimensional feature in the matrix V separately. The reason for dividing by the root sign d_k is to reduce the range of dot product, and the global semantic sharing is realized by the cyclic iterative computation of Attention Attention Mechanism model, and the output of each Attention Mechanism unit is represented as:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{5}$$

In the above equation, Q, K, V is three linear transformation matrices, and the value of d_k is determined by the K matrix dimension.

Multiple attention mechanism units form multi-headed attention, and the Q, K, V dimensions of d_{model} are mapped into d_k, d_k, d_v dimensions respectively by n linear transformations, and based on the mapped Q, K, V, the attention function is executed in parallel to output the values of the $h \times d_v$ dimensions, and then the final values are output after joining and mapping, and the formula is as follows:

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \cdots, head_n)W^o \tag{6}$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V), i = 1, 2, 3, \cdots, n \tag{7}$$

In Eq. (6) and Eq. (7): W_i^Q, W_i^K, W_i^V are all mapping matrices, $W_i^Q \in R^{d_{model} \times d_k}, W_i^K \in R^{d_{model} \times d_k}, W_i^V \in R^{d_{molel} \times d_v}$, $W^o \in R^{d_{model} \times hd_v}$, $d_{model} = 512$, and the number of polytopes $n = 8$.

The aggregation and optimization using the attention mechanism is mainly in two aspects:

(1) Node aggregation

Node aggregation is mainly realized by the information of the former nodes and edges in the sequence. Firstly, given a node feature C_i^t and a set of edge features $E_{C_i}^t = \{\forall_k e_{k,j}^t\}$, and then the attention mechanism is utilized to aggregate the current node C_i related to edge features $e_{k,j}$ corresponding to the former nodes feature C_k, the node aggregation function is formulated as follows:

$$e \to c : C_i^t = MultiAtt(c_i^t, \{\forall_k [c_k^t; e_{ki}^t]\}) \tag{8}$$

where t denotes the number of aggregations and $[c_k^t; e_{k,i}^t]$ denotes the links between nodes and edges.

(2) Aggregation of edges

In this paper, the aggregation of edges is realized by matching the complete input character sequence, and the function formulation is shown in equation (9):

$$c \to e : e_{b,f}^t = MultiAtt(e_{b,f}^t, C_{b,f}^t) \tag{9}$$

The compositional part can be known as the character embedding sequence $C \in R^{n \times d}$ and the potential word embedding sequence $E \in R^{m \times d}$, which denotes the initial node state as C^0, and the first time character embedding as the initial state of the edge E^0. In order to capture specialized long word-dependent features, a global relay node g^t is introduced to aggregate the features of each character and edge as follows:

$$c,e \rightarrow g : g^t = [g_c^t ; g_e^t] \tag{10}$$

$$g_c^t = MultiAtt(g^t, C_{1,n}^t)(3-13) \tag{11}$$

$$g_e^t = MultiAtt(g^t, \{\forall e^t \in \varepsilon\}) \tag{12}$$

II. A. 4) CRF decoding

Although the embedding layer and graph attention layer can learn the semantic relations of text context better, they are deficient in dealing with labeled sequences with long dependencies and multi-features, so based on the BMES sequence annotation method, the CRF decoding layer is utilized to learn the relational features of these labels.

The input of CRF named entity recognition model is a list of contextual feature vectors X(n) from Bert-Graph Attention model, and the output state sequence Y(n) is the entity labeling sequence. The CRF model can automatically learn the relational constraints of the current word during the training process and predict the globally optimal labeling sequence by combining the global probability of labeling sequence.

Given a list of context embedding vectors $X = (x_1, x_2, x_3, \cdots x_n)$, and a sequence of output labels $Y = (y_1, y_2, \cdots y_n)$, the probability maximizing sequence of output labels computes the formula:

$$P(X,Y) = \sum_{i=1}^{n} W_{y_i, y_{i+1}} + \sum_{i=1}^{n} N_{i, y_i} \tag{13}$$

where W is the state transfer matrix, dependent on the current positional state and the next positional state condition $w_{y_i, y_{i+1}}$, denoting the probability of labeling from the current y_i labeling type to y_{i+1}. N_i, y_i is the probability that the current word i is a label of type y_i. $P(X,Y)$ represents the probability of labeling the input sequence of sentences X as a sequence of labels Y, and the set of labeled sequences with the largest probability value $P(X,Y)$ is taken as the final sequence.

II. B. Civic content entity relationship extraction

This subsection focuses on how to extract relationships between entities based on the entity acquisition in the previous section. Due to the lack of structured data in the content of Civic and Political Education, this project mainly utilizes unstructured data and semi-structured data related to Civic and Political Education as the corpus for relationship extraction. The semi-structured data are mainly the information box data of the entries of Civic and Political Figures, Political Organizations and Political Conferences in the Chinese encyclopedia web pages, which are mainly used to extract the relationship triples by crawler technology. Unstructured data are the plain text data related to the contents of Civic and political education obtained from websites such as “Learning Power”, “Communist Party of China”, and “Zhonggong Education Current Affairs Channel”, which are mainly extracted using an attention-based fusion mechanism. This part of the data mainly utilizes an entity-relationship extraction method based on the attention mechanism fused with syntactic analysis to obtain the Civic and Political Relationship Triad.

II. B. 1) Relational extraction of semi-structured defeat data

Semi-structured data mainly exists in various encyclopedia sites, the semi-structured data of Civics education content in this paper mainly comes from the information box data of Civics entries in Baidu Encyclopedia, due to the uniformity of the storage rules of the information box data, it is only necessary to filter out the Civics entries that need to be extracted, and then build the data wrapper according to the rules of the information box data storage in the encyclopedia page, and then the semi-structured data can be extracted, which can obtain the “entity-relationship-entity” form of the knowledge triad of Civics. Then we can obtain the “entity-relationship-entity” form of the Civics knowledge triad. Currently, the mainstream webpage information crawling framework is BeautifulSoup, which is also used in this paper as a tool for extracting the relationship between Civics and Politics in Baidu Encyclopedia.

II. B. 2) Relational extraction of unstructured defeat evidence

Unstructured text data in the field of Civics is characterized by diversity, ambiguity, varying lengths and noise interference, in this regard, this paper proposes an entity-relationship extraction method based on the attention mechanism fused with syntactic analysis for Civics text for relationship extraction.

(1) Word Representation Input Layer

In Chinese, a word does not only contain one character, but may consist of multiple characters to form a word, and the characters contain rich internal structural information between them, so this paper uses Chinese word vectors trained by Word2Vec tool. The difference between it and word vectors is that word vectors use a vector to represent a word, while word vectors use a vector to represent a character.

Given a sentence $S=\{x_1,x_2,\cdots,x_n\}$, each word x_i in the original data is converted into a character vector e_i. This is done by first converting each word in the sentence S into a matrix of word vectors $W^{word}\in R^{d^w\times V}$, where V is the fixed-size lexicon and d^w is the size of the word embedding. $W_i^{word}\in R^{d^w\times V}$ represents the word vector of the i th word in the dictionary. The word x_i is mapped to the vector e_i using matrix-vector multiplication as shown in equation (14):

$$e_i=W^{word}\times v^i \tag{14}$$

where v^i is the one-hot representation of a dictionary of size v, with a value of 1 at the subscript e_i and 0 elsewhere, thus converting the Chinese utterance into a vector $emb_s=\{e_1,e_2,\cdots,e_n\}$, which serves as an input to the subsequent model.

(2) Attention mechanism

In order to improve the performance of the model used to deal with entity relationship classification in this paper, an attention mechanism is added to the model to automatically assign weights to the input vectors to obtain valid information and reduce the influence of noisy data. This section proposes to add a word-level based attention mechanism and a sentence-level based attention mechanism to the bidirectional GRU neural network model to improve the accuracy of the Civic Education content relationship extraction task.

The weights in the attention mechanism are usually computed in the following way:

$$a_t=\frac{\exp(f(m_t,n))}{\sum_{k=1}^{l}\exp(f(m_k,n))} \tag{15}$$

where a_t is the weight of the vector m_t computed automatically in the attention mechanism. f is a function that associates the vector m_t for which weights need to be computed with the vector n corresponding to the factors affecting the weights. l denotes the number of all vectors that need to be assigned weights. The a_t is normalized using softmax such that the weights of all vectors add up to 1. Then:

$$f(m_t,n)=v_a^T\tanh(W_a m_t+U_a) \tag{16}$$

where v_a is the weight vector and W_a and U_a are the weight matrices.

Let the output vector matrix $H=\{h_1,h,\cdots,h_n\}$ generated by the two-way GRU neural network layer, where n is the sentence length. The representation γ of the sentence is formed by the weighted sum of the following output vectors:

$$M=\tanh(H) \tag{17}$$

$$\alpha=softmax(W^T M) \tag{18}$$

$$\gamma=H\alpha^n \tag{19}$$

where $H\grave{o}R^{d^{w\times n}},M\in R^{d^{w\times n}}$, the tanh function is used to transform the original vectors between [-1, 1]. d^w is the dimension of the word vector, W is a trained parameter vector, W^T is its transpose, and α has size n.

$$h^*=\tanh(\gamma) \tag{20}$$

where W is the weight vector of the attention mechanism layer, a parameter to be trained. h^* is used as the sentence representation after weighted summation by the attention mechanism layer.

(3) GRU model based on attention mechanism fusion syntactic analysis

a) Dependent syntactic analysis

In order to be more clear about the syntactic structure of a sentence, it is necessary to analyze the dependency relationship between the components in that sentence, so syntactic analysis according to law is proposed. The LTP natural language processing tool of Harbin Institute of Technology can better analyze the dependency syntactic relationship among the components of Chinese sentences.

b) GRU Model

GRU model is obtained from LSTM after simplification, the model structure is simpler than LSTM, and also maximizes the same effect with LSTM model, the parameter of GRU neural network is reduced by 1/3 compared with LSTM, which is not easy to produce overfitting, and the convergence time is also faster than LSTM model, and the number of iterations is less than LSTM model. Then:

$$r_t = \sigma(W_r)\cdot[h_{t-1}, x_t]) \tag{21}$$

$$z_t = \sigma(W_z)\cdot[h_{t-1}, x_t]) \tag{22}$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}})\cdot[r_t * h_{t-1}, x_t]) \tag{23}$$

$$h_t = (1-z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{24}$$

$$y_t = \sigma(W_0 \cdot h_t) \tag{25}$$

where x_t denotes the input at the current moment, h_{t-1} denotes the output at the previous moment, $W_r, W_z, W_{\tilde{h}}, W_0$ are the corresponding weight matrices, z_t and r_t are the update and reset gates, respectively, [] denotes the concatenation of matrices, and * denotes the matrix multiplication.

In the bi-directional GRU model consists of both front and back networks together, the forward direction takes input from the front end of the sentence, and the backward direction takes input from the back end of the sentence into the network, and then splices the information of the front and back directions according to the position to get the vector of feature representations of the sentence.The output of the BiGRU model is represented by Eq. (26):

$$h_i = [\overrightarrow{h_t} \oplus \overleftarrow{h_t}] \tag{26}$$

The output of the model uses the results of element-based forward and backward splicing.

Steps of the BiGRU neural network model for Attention Mechanism Fusion Syntactic Analysis:

The first step is to input an utterance and label the sentence with split words.

In the second step, the trained distributed word vectors are inputted into the bidirectional GRU network to obtain the deep semantic information of the sentence, the attention mechanism is introduced to pay attention to the important information in the sentence, and the influencing factors affecting the entity-relationship extraction task are automatically obtained by the attention mechanism automatically updating the characteristics of the sentence weights.

Finally, the processed sentence weight vectors are obtained and then the entity relations are classified using softmax classifier.

III. Experimental design and discussion of results

III. A. Entity Naming Recognition Analysis

III. A. 1) Experimental design

The experimental data is the self-constructed Civic Education Content Knowledge Entity dataset made by manual and semi-automatic tagging in this paper, with a total of 12,520 corpora. The pre-processed corpus is randomly disrupted and divided into three datasets according to 8:1:1 for training, validation and testing.

III. A. 2) Comparison of Entity Recognition Models

On the self-constructed dataset, HMM, CRF, BiLSTM, BERT-softmax, BERT-CRF were used to compare the experiments with the Bert-Graph Attention-CRF model in this paper. The comparison of the recognition results of Civic objective existence entities and Civic logical concept entities are shown in Figures 1 and 2. In the experimental training, Adam is chosen as the optimizer of the model, with a learning rate of 0.001, a batch size of 64, 100 epochs, and a dropout of 0.25. In order to prevent the model from overfitting, this paper stops the model from training by setting the model to stop training when its loss on the validation set will not drop for 20 epochs in a row. In Civics objective existence entity recognition, the F1 score of this paper's model is 96.23%, which is 1.31%~42.70% higher than the comparison methods. In the Civics Logical Conceptual Entity Recognition, the F1 score of this paper's model is 88.46%, which is 1.29%~41.65% higher than the other methods.

The experimental results show that the BiLSTM-CRF model has a better recognition effect than traditional named entity recognition models such as HMM, CRF, BiLSTM, etc. Therefore, this paper chooses to use the Bert-Graph Attention-CRF model as a model for named entity recognition of the content of Civics education.

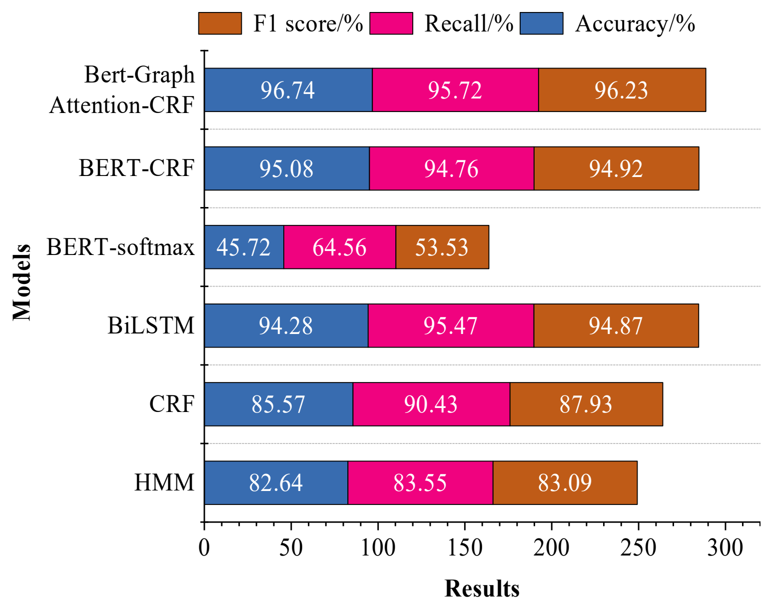

Figure 1: The comparison of the recognition results of the objective existence entity

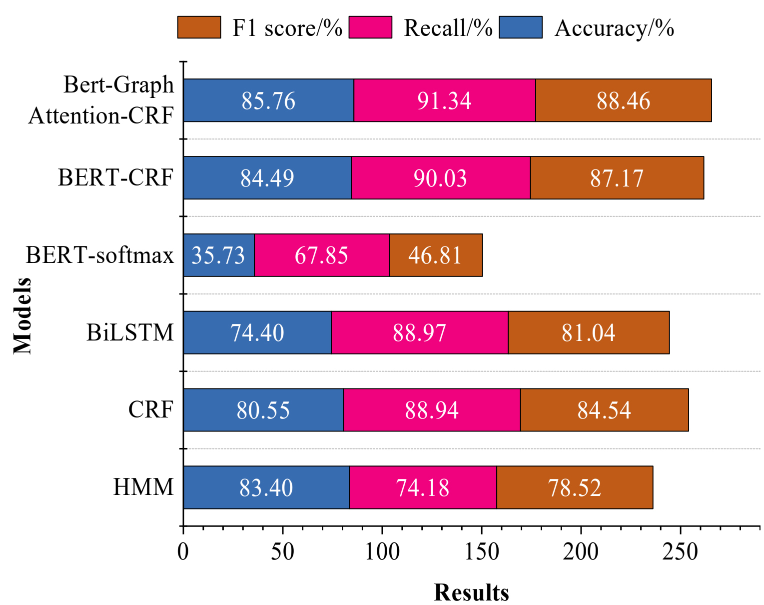

Figure 2: The comparison of the recognition results of the logical conceptual entity

III. A. 3) Selection of the number of nodes

After the comparison test, it can be concluded that the Bert-Graph Attention-CRF model is significantly better than the other five models, but the choice of the number of nodes in the Bert-Graph Attention-CRF model is crucial. Too many nodes will lead to a complex network and low generalization ability. Too few nodes will lead to weaker model learning ability and inaccurate results due to decreased feature capture ability, so this experiment is set up to explore the optimal number of nodes. In this experiment, the number of nodes of Bert-Graph Attention-CRF model is set to 25, 50, 75, 100, 125, 150 for the experiment to judge the optimal number of nodes of the model. Then the optimal number of nodes is used for entity recognition as judged by the experimental evaluation index. The experimental results for different nodes are shown in Figure 3. When the nodes of Bert-Graph Attention-CRF model are set to 125, the model has the highest accuracy, recall and F1 value, which are 91.96%, 92.35% and 92.15%, respectively. Therefore, the Bert-Graph Attention-CRF model is selected for entity naming recognition and its node number is set to 125.

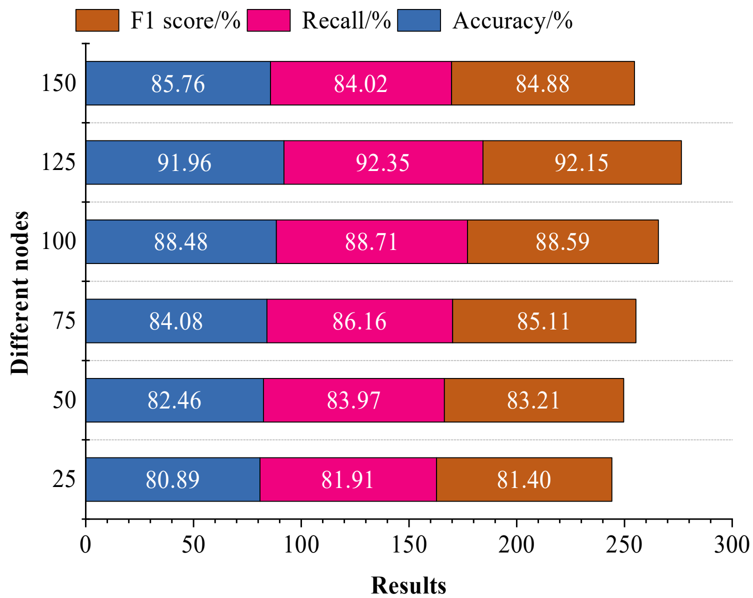

Figure 3: Experimental results of different nodes

III. A. 4) Entity acquisition results

After the previous section, it is known that the Bert-Graph Attention-CRF model works best when 125 nodes are selected, so this model is used for entity recognition, and the acquired entities of Civic Education content are shown in Table 1. A total of 5246 entities in the field of Civics and Politics were acquired, and 746 valid entities were retained after manual de-duplication and merging.

Table 1: Results of entity recognition

Category of entities	The number of correct entities obtained	Quantity after manual
Political figure	1727	391
Political activity	594	105
Political organization	253	64
Site	2672	186
Total	5246	746

III. B. Entity Relationship Extraction Analysis

The corpus of this experiment comes from the combination of Scrap crawler technology framework and manual extraction, 1608 textual Civics text data in the form of text from a number of Civics-related websites and Baidu Encyclopedia were screened after utterance splitting, and finally 1500 utterances were selected as experimental materials. The length of these statements does not exceed 30 characters, and they are highly relevant and relatively regular to the information of Civic and Political Affairs. These utterances were categorized into six groups A, B, C, D, E and F for experimentation, and each group contained 250 utterances each for comparative study. These corpora have been manually labeled before the experiment. Eventually, the acquired lexicon consisting of Civic entities was deep learned using the BiGRU model of Attention Mechanism Fusion Syntactic Analysis, and the Civic entity relations were successfully targeted and extracted.

Experiments on the acquisition of relations of ideological and political entities were carried out on six corpora, A, B, C, D, E, and F, and the results of the experiments were counted, and the results of the six groups of relation extraction are shown in Fig. 4. The accuracy, recall, and F1 value of Group E showed the best results compared with Groups A, B, C, D, and F, and the percentage gap between the three evaluation indexes of Groups A, B, C, D, and F was no more than 5.26%, for this reason, it can be stated that the present experiment was conducted on the existing high-quality topic corpus and with pragmatically constrained entities. Based on the above findings, this time, the BiGRU model of fusing syntactic analysis through the attention mechanism will be used to extract entity relations for the remaining part.

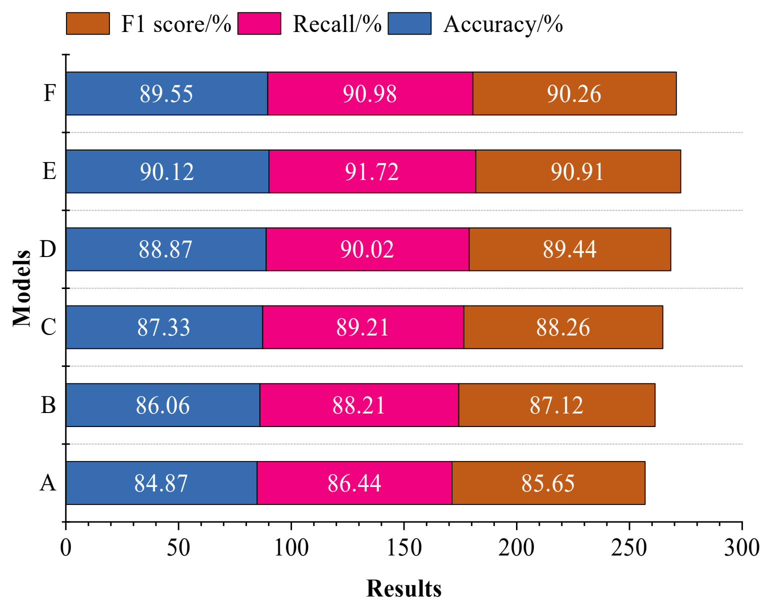

Figure 4: Six sets of relationship extraction results

IV. Intelligent organization and management of civic education content

This chapter explores the method of intelligent organization and management of Civic Education content based on the knowledge extraction method of Civic Education content, and constructs the Civic Education content knowledge graph.

IV. A. Building the process

This topic is based on the characteristics of the topic of Civics and the design principle of knowledge graph, and gradually build a visualization system based on knowledge graph. The construction process of Civic and Political Education Content Knowledge Graph is shown in Fig. 5, including corpus construction, ontology construction, schema layer construction, data layer construction, knowledge storage and visualization steps.

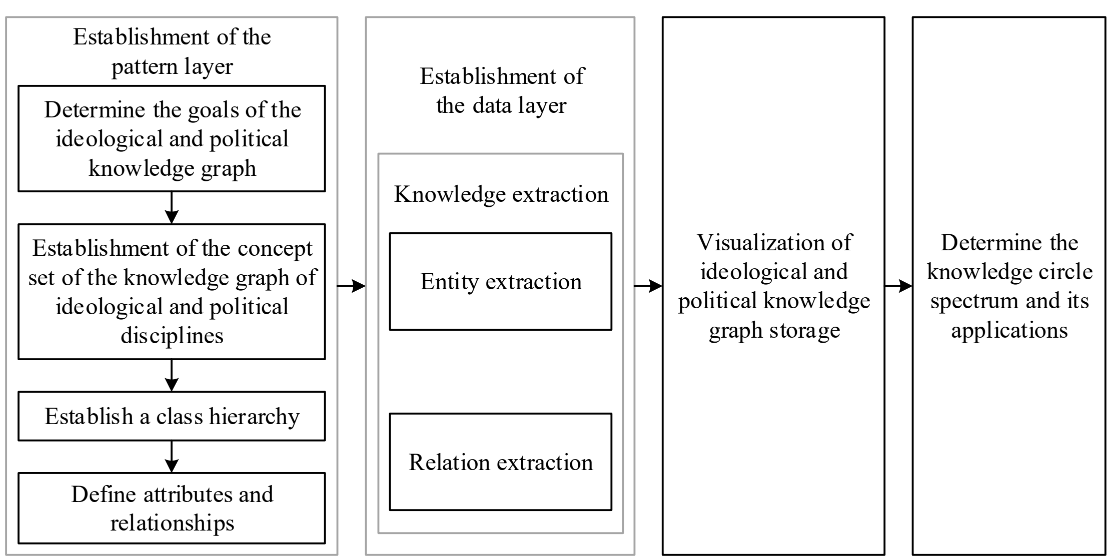

Figure 5: The construction process of the knowledge map of the ideological and political education

IV. B. Civic Education Knowledge Corpus Construction

(1) Data sources

In order to ensure the quality of the corpus, the data constructed in this paper comes from the following three main sources: policy documents and university cultivation programs, Civics course materials and course syllabi, social media and the Internet.

(2) Civics Text Corpus Labeling

The field of Civics courses is mostly semi-structured and unstructured data, and the raw data text of the experimental corpus for the Civics course named entity recognition task is extracted manually using a combination of Scrapy crawler technology framework and manual extraction, and is annotated on websites such as “Learning Power”, “Xinhua.com Xinhua Civics” and Baidu Encyclopedia. The unstructured data in the field of Civics and Political Science in the form of plain text is crawled from the web pages of “Learning Power”, “Xinhua.com Xinhua Civics and Baidu Encyclopedia” and other websites.

To address the special semantic difficulty phenomenon of Civics education content knowledge, the four-bit sequence annotation method (BMES) was used to annotate the corpus of the Civics education content named entity recognition task corpus. The following six categories were selected: name of the civic and political figures NAME, content of political activities CONT, title attribute of the civic and political figures TITLE, profession of the civic and political figures PRO, name of the political organizations and institutions ORG, and location of the political events LOC.

IV. C. Ontology construction

(1) Ontology construction tool

Protégé was chosen for this paperThis paper chose an object-oriented generative tool for the development of domain ontologies.Protégé is a Java language-based software especially designed for creating and editing ontologies.

(2) Ontology construction methods

At present, the common methods of ontology modeling include the following five: assessment method (Tov), skeleton method (Sensus), Methontology method, five-step method, seven-step method. According to the characteristics of the content of Civic Education, the seven-step method, which is widely used for ontology construction in the field of education, is selected.

IV. D. Schema Layer Construction

(1) Ontology Concept Determination

The purpose of this paper is to expand the basis for the application and visualization of the existing knowledge mapping of the content of Civic and Political Education, and to supplement new knowledge with the nature of Civic and Political, so as to make the structure of Civic and Political knowledge richer. By organizing the collected data and studying and analyzing the Civics curriculum, the Civics knowledge is divided into knowledge points, knowledge blocks and knowledge systems according to different levels of size. This study focuses on the two main clues of “concept composition” and “concept relationship”. Concepts can be composed of knowledge points, knowledge blocks and knowledge systems. In order to determine the knowledge concepts of Civics and Political Science curriculum, on this basis, a knowledge-based concept nucleation method is proposed and semantic extractions, such as instances, attributes, etc., are carried out.

(2) Determination of conceptual hierarchical relationship

In the content of Civic Education, there are two main types of concept hierarchy relationships. The first type is the subset relationship, in which concepts are hierarchically categorized according to the discourse structure in the standards and textbooks. The second type is conceptual relations, which are based on the organization of knowledge and organize concepts in a hierarchical order.

(3) Definition of Object Attributes and Relationships

Through the organization and analysis, the object attribute relationships of the Civics knowledge points are summarized: inclusion relationship, synonymous relationship, antecedent relationship, successor relationship, brotherhood relationship, and antagonism relationship.

IV. E. Data layer construction

The data layer of Civic Education Curriculum Knowledge Graph refers to modeling and organizing the knowledge in the field of Civic Education Curriculum to form a graph with entities, attributes and relationships as the basic elements for knowledge query, reasoning and application. Through the research on knowledge extraction of Civic and Political Education Content Knowledge Graph in the second chapter of this paper, after entity extraction, entity relationship extraction, knowledge fusion and knowledge storage are carried out step by step. Realize the construction of the data layer of the knowledge map of the content of Civic and Political Education.

IV. F. Knowledge storage and visualization

Neo4j, as a common graph database, uses Cypher Query Language (CQL), and in this paper, Neo4j graph database is chosen for knowledge storage. The file storing the ontology knowledge is converted into CSV format and saved to the import folder of Neo4j. The CSV files can be introduced in batches by LoadCSV command in Neo4j's Cypher syntax. Finally, to patch nodes that cannot be imported in bulk, use a Cypher statement to manually create them and their relationships.

V. Conclusion

In this paper, a complete set of technical solutions and implementation frameworks are formed through in-depth research on the intelligent organization and management method of Civic and Political Education content based on knowledge graph. Experimental validation shows that when the number of nodes of the Bert-Graph Attention-CRF model is set to 125, it achieves the best performance in the task of Civic and Political Entity Recognition, with an accuracy rate of 91.96%, a recall rate of 92.35%, and an F1 value of 92.15%. The relational extraction experiments choose 1500 high-quality statements for validation, and the results of the six groups of experiments A, B, C, D, E and F show that the model has good stability and generalization ability. The constructed knowledge map of Civic and Political Education contains a conceptual system at multiple levels, effectively organizing the core knowledge elements in the field of Civic and Political Education.

The successful application of knowledge mapping technology in the field of Civic and Political Education provides a new technical path for the development of educational informatization. The proposed entity recognition and relationship extraction methods can automatically process large-scale Civics text data, significantly reduce the manual annotation cost, and improve the efficiency of knowledge organization. The knowledge storage and visualization scheme based on Neo4j graph database realizes the structured management and intuitive display of Civic and Political education content, which facilitates knowledge retrieval and learning for teachers and students. This research can provide reference and reference for the construction of knowledge graphs in other fields, promote the in-depth application of artificial intelligence technology in the field of education, and promote the modernization and development of Civic and Political Education.

References

[1] Huang, P. (2024). Research on Personalized Ideological and Political Education Content Distribution System Based on Intelligent Algorithms. International Journal of High Speed Electronics and Systems, 2540151.

[2] Qing, Z. (2024). Innovating Content and Methods of Ideological and Political Education in the Context of the New Era. International Journal of Education and Humanities, 4(4), 479-487.

[3] Kosherbayeva, A. N., Dementieva, N. G., Sansyzbayeva, D. B., & Kosherbayeva, G. N. (2023). Effective components in the management structure of an educational organization. Bulletin of LN Gumilyov Eurasian National University. Pedagogy. Psychology. Sociology series., 145(4), 145-161.

[4] Susnea, I., Vasiliu, G., & Mitu, D. E. (2013). Enabling Self-Organization of the Educational Content in Ad Hoc Learning Networks. Studies in Informatics and Control, 22(2), 143-152.

[5] Shavdirov, S. (2025, February). Method of organization of classes in higher education institutions using flipped classroom technology. In AIP Conference Proceedings (Vol. 3268, No. 1, p. 070035). AIP Publishing LLC.

[6] Sherayzina, R. M., Monakhova, L. Y., & Maron, A. E. (2020). Educational content management in information and learning environments. European Proceedings of Social and Behavioural Sciences.

[7] Vaganova, O. I. (2019). Organization of practical classes in a higher educational institution using modern educational technologies. Amazonia Investiga, 8(23), 81-86.

[8] Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. Procedia computer science, 136, 16-24.

[9] Trivedi, R., Dai, H., Wang, Y., & Song, L. (2017, July). Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In international conference on machine learning (pp. 3462-3471). PMLR.

[10] Zhang, Y., Dai, H., Kozareva, Z., Smola, A., & Song, L. (2018, April). Variational reasoning for question answering with knowledge graph. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).

[11] Yang, D., Liu, S., Fu, H., & Shen, J. (2024, September). Research and Practice on the Construction of Course Ideological and Political Education Based on Knowledge Graphs and Large Language Models. In Proceedings of the 2024 International Symposium on Artificial Intelligence for Education (pp. 193-198).

[12] Zhu, G., & Tao, T. (2022, December). Application of Knowledge Graph in Ideological and Political Teaching. In 2nd International Conference on Internet, Education and Information Technology (IEIT 2022) (pp. 858-864). Atlantis Press.