Design of a knowledge graph-based question and answer system for teaching assistance

Ping Feng
PhD student
Changchun University
Changchun, China
fengping@ccu.edu.cn
ORCID 0000-0003-4865-1454

Xu Siqi
Magister
Changchun University
Changchun, China
2786265330@qq.com
ORCID 0000-0000-0000-0000

Yingying Wang
Magister
Changchun University
Changchun, China
2411207712@qq.com
ORCID 0000-0000-0000-0000

Received 05.12.2024
Accepted 26.01.2025
Published 15.02.2025

UDC 004.421.2
DOI 10.25726/b2852-1767-1381-l
EDN DYIUOV
VAK 5.8.2. Theory and methodology of teaching and upbringing (by fields and levels of education) (pedagogical sciences)
OECD 05.03.HB. EDUCATION, SCIENTIFIC DISCIPLINES

Abstract

The aim of this paper is to construct a knowledge graph-based teaching-aided Q&A system that focuses on solving students' preview needs in a course on classical Chinese masterpieces. By combining knowledge mapping with multiple resources, the system is able to enrich the form of traditional cultural communication and provide targeted Q&A services. The system aims to help students prepare for the course in advance, get timely answers to questions, reduce the burden of teachers, improve the efficiency of classroom teaching and the quality of student learning, and explore ways to disseminate traditional Chinese culture.

Keywords

knowledge mapping; question and answer system; classical Chinese masterpieces; traditional culture dissemination.

This article is supported by the following project funds： Jilin Province's 2023 Higher Education Association Research Project (Research on the Integration of Artificial Intelligence and the Inheritance of Excellent Traditional Chinese Culture). Project Approval Number: JGJX2023D397. Jilin Province's 2023 Vocational Education and Adult Education Teaching Reform Research Project (Research on the Deep

Integration of Artificial Intelligence Technology into the Entire Process of Teaching and Management Practice) Project Approval Number: 2023ZCY352.

Introduction

An intelligent Q&A system is a system that can automatically retrieve relevant information from a knowledge base and generate answers based on questions input by the user. With the continuous development of natural language processing technology, the performance of intelligent Q&A systems has been significantly improved. In the field of education, intelligent Q&A systems have been used to assist teaching and provide personalized learning support for students. In this paper, we will design and implement an intelligent Q&A system for the questions of a course on classical Chinese masterpieces to meet the students' pre-study needs (An, 2023).

Under the background of today's booming network culture, classical Chinese masterpieces, as an important part of Chinese culture, are facing new opportunities and challenges in their dissemination and learning methods. In view of the fact that students often encounter a lot of questions when previewing a course on classical Chinese masterpieces, and that these questions often appear in the form of a large number of repetitive questions in the classroom, this paper proposes an innovative teaching aid, a question and answer system based on knowledge mapping (Cauteruccio, 2023). The system aims to build an efficient and convenient Q&A platform by integrating and parsing the relevant knowledge of classical Chinese masterpieces in order to meet the information needs of students in the pre-study process. The research in this paper not only helps to enrich the form of dissemination of traditional culture, but also effectively improves students' previewing effect, reduces teachers' repetitive Q&A burden, and thus promotes the in-depth teaching and learning of classical Chinese masterpieces in the course (Zhiqiang, 2023).

Knowledge graph is a structured knowledge representation that presents entities, concepts and their relationships in the form of graphs. In recent years, knowledge graphs have been widely used in intelligent Q&A, recommender systems, semantic search and other fields. In the field of education, knowledge graphs are also used to build subject knowledge bases to help students better understand and master knowledge. In this paper, we will draw on these successful experiences and utilize knowledge mapping technology to construct the knowledge system of Chinese classical masterpieces to provide accurate and comprehensive knowledge support for the Q&A system (Chen, 2022).

Materials and research methods

As an important carrier of Chinese culture, the digital dissemination of classical Chinese masterpieces has become a hot spot of current research. Through digitization, classical Chinese masterpieces have been widely disseminated globally, providing more people with the opportunity to contact and understand Chinese culture. In this paper, we will make full use of these digitized resources and integrate them into the knowledge graph to provide a rich source of information for the Q&A system (Chen, 2022).

Teaching support systems are an important means of improving the quality and efficiency of teaching. In recent years, with the continuous development of information technology, a variety of teaching aid systems have emerged to provide teachers and students with convenient learning and teaching support. In this paper, we will draw on the successful experience of these systems and combine them with the characteristics of a course on classical Chinese masterpieces to design and implement a knowledge graph-based teaching aid question-and-answer system in order to improve students' previewing effect and teachers' teaching efficiency.

This system obtains structured and unstructured data from books related to the introduction of masterpieces, Baidu Encyclopedia, Zhihu and other websites for constructing the knowledge graph. Different processing methods are used for data with different degrees of structuredness, and the data are transformed into the form of ternary groups, and finally stored in the Neo4j database (Qian, 2021). The flowchart for constructing the knowledge graph of Chinese classical masterpieces is shown in 1 below.

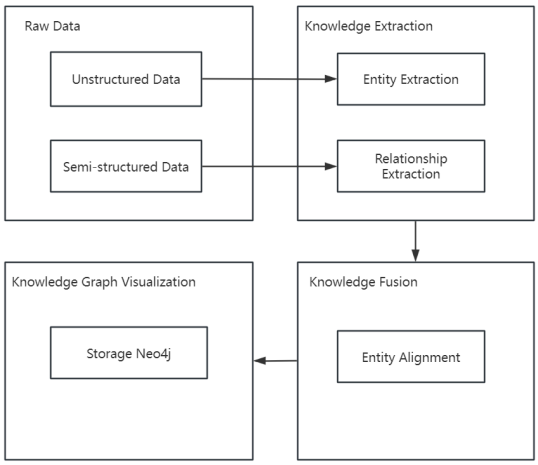

Figure 1. Flow of constructing knowledge map of Chinese classical masterpieces.

Classical Chinese masterpieces are an important part of Chinese classical literature. There are various genres such as poetry, prose, novels, dramas, as well as words, songs, fugues, folk myths and legends. First of all, writings related to classical Chinese masterpieces are one of the important sources for constructing the knowledge map of classical Chinese masterpieces. In order to enable students to have a standardized and systematic understanding of classical Chinese masterpieces. Relevant scholars will write relevant writings on the basis of history and according to the rigor of what they have learned to introduce some relevant literary knowledge to the readers. The contents of the writings are usually hierarchical and help readers to systematically understand the relevant knowledge in the field, so these writings are an indispensable part of the preparation of the knowledge map (Do, 2021).

Results and discussion

However, for some specific writings, the contents told are limited and it is impossible to cover all the contents of classical Chinese literature, but combining multiple writings together is too much repetition, and at the same time, it is not possible to ensure the completeness of the data. Therefore, in this paper, in addition to selecting some representative masterpieces, the knowledge graph is expanded by combining the relevant contents of Baidu Encyclopedia, blogs and other fields, which makes the constructed graph more complete.

Web crawling techniques can be used to quickly obtain information from large amounts of semi-structured data. In this paper, we use the Scrapy framework for data acquisition. Scrapy is a high-level, fast screen information crawling and web crawling framework for Python (Rúben, 2022). Scrapy also provides many convenient features, such as auto-retry, dynamic User-Agent, distributed crawling, and anti-crawling mechanism, which can help users to more efficiently accomplish data crawling and processing. It is worth noting that you need to comply with relevant laws and regulations and ethical norms and respect the privacy policy of the Web site when performing Web crawling. In order to avoid being blocked by the anti-crawler mechanism, it is necessary to take some measures, such as restricting the access frequency, using proxy IPs, and simulating user behavior. The architecture diagram is shown in Figure 2 below.

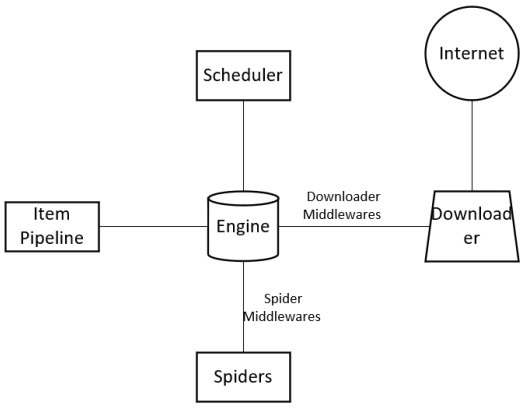

Figure 2. Scrapy architecture diagram.

Data preprocessing of the crawled unstructured data, with period, question mark and exclamation mark as the end mark, using Stanford Chinese Toolkit for the «Introduction to Classical Chinese Masterpieces» for the processing of sentence and word division, using Stanford Pos tagger for the lexical annotation, and finally generating embedded files to be saved in the form of .pkl. Test data, training data and validation data for constructing named entity recognition model experiments.

The BiLSTM-Highway-Hypergraph model proposed in this paper is mainly applied to the recognition of nested entities, and the model adopts the spanning representation for entity recognition on the basis of the hypergraph model, which quickly and accurately recognizes all possible entity boundaries, and solves the problem of boundary ambiguity to some extent. Firstly, in the input layer, the feature vectors of the words trained in advance are used as inputs, combined with the features specific to the hypergraph to effectively obtain the relevant information of the potential words. Secondly, after obtaining the high-dimensional features, the input layer feature representations are fed into the BiLSTM model for feature extraction and generation of effective contextual information representations. Then, Highway neural network is used to obtain the word representation of interest in the sentence and finally the model combines the relevant features of the hypergraph to generate all the possible entities and output the recognition results. The structure of the model is shown in Figure 3 below.

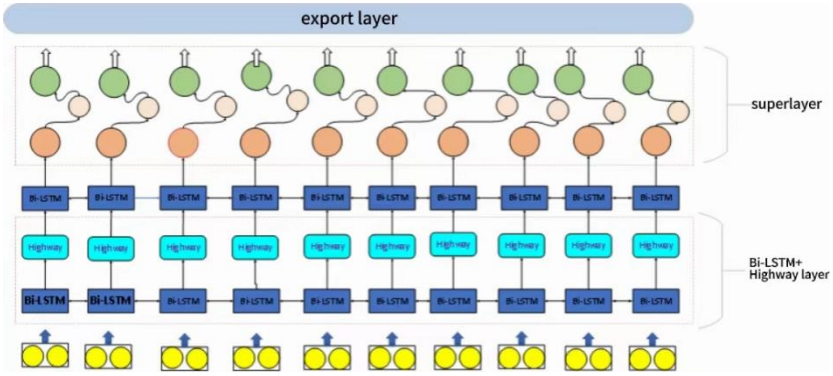

Figure 3. Model architecture diagram of BiLSTM-Highway-Hypergraph.

The core part of the model is to use the hypergraph model to solve the nested entity recognition problem. The first layer is the word granularity embedding layer, which obtains the feature vector representation of words; the second layer, Bi-LSTM+ Highway coding layer, captures word context information to form word-level features; the third Bi-LSTM layer, which generates fragment-level features by incoming information from the previous layer; the fourth layer is the hypergraph layer, which computes the fractional function of the hyperedge to form different combinations of fragments to recognize nested entities (Feiran, 2022).

Much of the data crawled from web pages will have some quality issues such as attribute name diversity, entity ambiguity, etc. The collected data is cleaned and processed for data fusion. In this paper, Synonyms, a

proximity toolkit, is used to determine the similarity of different source information. Synonyms based on Word2vec works by first determining whether the attributes of attractions from two different sources are similar. If they are not similar, the book title and related attributes are added. If they are similar, further determine the similarity of the attributes, and if the attributes are not similar, add the attributes.

Analyzing the data from the construction of the knowledge map of Chinese classical masterpieces, the following features are derived (Li, 2022):

1) For the same entity from different sources, different web pages have similarities in the main content of its description. For example, for the same author's introduction, different webpages have different degrees of accuracy in the introduction of their dynasties.

2) There are more cases of missing attributes. For example, not all authors consist of attributes such as word and number.

3) The number of entity pairs to be matched is not high. The article carefully groups authors and famous books based on the dynasty in which they are located, so there will not be too many alternatives when matching entities.

Knowledge graph in the data are stored in the form of ternary groups, the data will be fused, the form of its processing, in the form of «attribute-attribute value» of the form of relations, with the data will be described as (entity, attribute, attribute value) in the form of storage form.

Neo4j has a specialized Cypher language for querying, adding, deleting and changing data. In this paper, in combination with the consideration of the number of nodes and relations of the knowledge graph and the construction efficiency, the data processed earlier are deposited into the Neo4j database in the form of ternary groups (Qian, 2021). All the nodes and relationships into the Neo4j graph database, the construction of the knowledge graph of classical Chinese masterpieces is completed, in which the nodes of different colors on behalf of the different types of entities, you can also set their own color to distinguish. Figure 4 below shows the overall display of the completed knowledge graph and some of the diagrams.

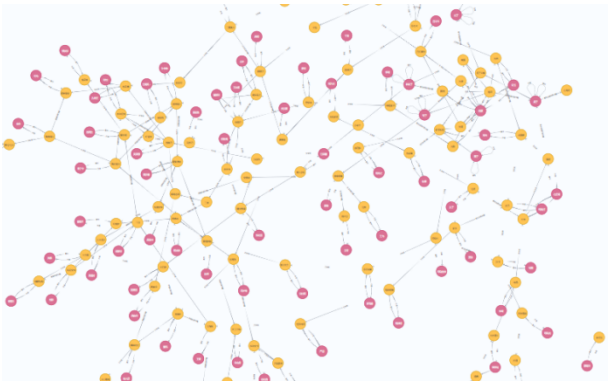

Figure 4. Display of knowledge mapping of Chinese classical masterpieces.

The main flow of the Q&A system framework designed in this paper is shown in Figure 5 below, the core of which is mainly to convert the user's question into an information query in natural language, and then match the answer from the database and return it to the user.

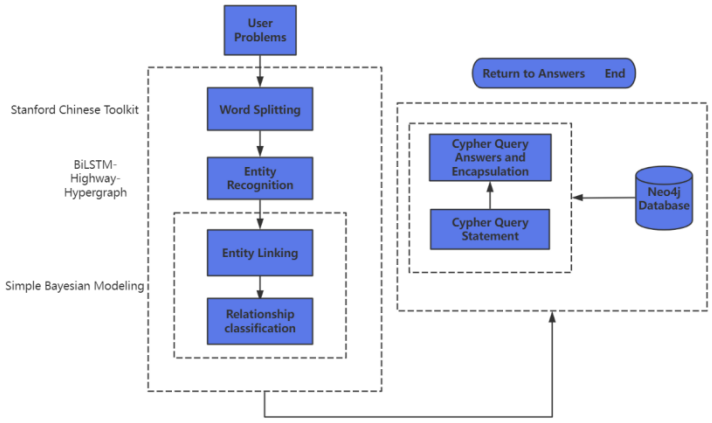

Figure 5. Q&A System architecture design diagram.

The entity recognition algorithm in this module uses BiLSTM-Highway-Hypergraph model for recognition. The specific recognition process is shown in Figure 6 below.

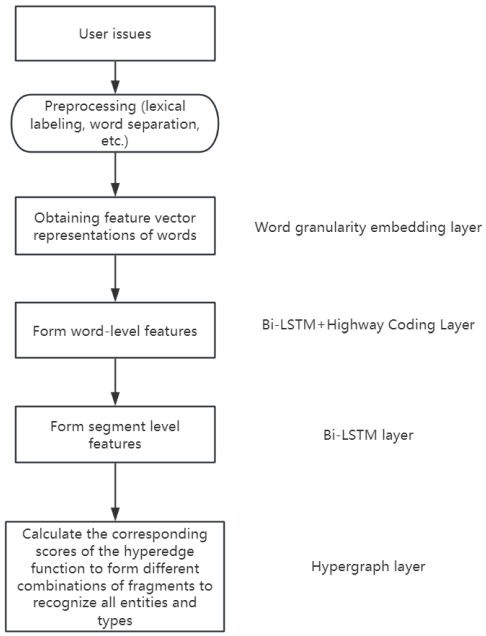

Figure 6. Entity recognition flowchart.

The first layer is the word granularity embedding layer, in which the user problem sentence is preprocessed, segmented, etc., as an input to this layer, where a feature vector representation of the word will be obtained; the second Bi-LSTM+ Highway coding layer, where the word with the obtained feature vector representation will capture the word contextual information to form the word level features, which will further serve as a data input to the next layer; the third Bi- LSTM layer, which generates segment level features by the incoming information from the previous layer; and the fourth layer is the hypergraph layer, which computes the fractional function of the hyper edges, which results in the formation of different combinations of segments, thus recognizing all the entities and types of entities (Hou, 2022).

Since there are fewer question-and-answer datasets about Chinese classical masterpieces, the relational dataset used in this experiment needs to be labeled by itself. The results of the interviews indicate that the users' questions about Chinese classical masterpieces mainly include eight types, who is the author of a certain book; what is the main content of a certain book; what are the main characters of the book; the influence of a certain book on the later generations, and the status of the literary world, and so on. The details are shown in Table 1 below. The source of the dataset is mainly the results of interviews and related searches on the

Internet, e.g., when querying what is the main story of Tian Gong Kai Wu, it can be expressed as «the main content of Tian Gong Kai Wu», «a brief introduction to Tian Gong Kai Wu» and so on. The training set mainly uses question templates, which contain different types of questions.

Table 1. List of question categories

№	Form	Example of a training set	Test set example
1	Who is the author of a particular book	Who is the author of a book	Who wrote the «Notes on the Weeping Cottage»?
2	What a particular book is about	What's the Info on a book?	What's The Wizard of Oz about?
3	What are the main characters of a particular book	Who is the main character of a certain book	Who is the main character of «Dream of the Red Chamber»?
4	The impact of a particular book on subsequent generations	What is the rate of a certain book	Later Evaluations of the Book of Pipa
5	What was the context of the time when a particular book was written	Background specifics of a book	The background of «Sinister Souls»
6	Where can I find some other relevant information about this book	What is the related of a particular book	Other relevant information on Menzies
7	When was the book written?	When is the time for a certain book	When was Lui's Spring and Autumn written?
8	Relationships between an author and a book or other authors	Relationships between an author and a book or other authors	What is the relationship between the Book of Admonition and Li Si?

After creating a good dataset according to the above method, the corresponding entities of the user are recognized after BiLSTM-Highway-Hypergraph model, and then the question matching template is designed to recognize the user's intention; the user's intention template dictionary is designed, and the mapping answer template dictionary is designed.

According to the designed template type, combined with Cypher statement rules, the following statement templates are constructed:

1) «Who is the author of a certain book» Cypher query statement match (p:Person{«book name»}) – (author) return p,author.

2) «What a book is about» Cypher query statement match (p:book{«book name»}) – (info) return p,info.

3) «What are the main characters of a particular book» Cypher query statement match (p:Person{«book name»}) – (actor) return p,actor.

4) «The impact that a particular book has had on subsequent generations» Cypher query statement match (p:book{«book name»}) – (rate) return p,rate.

5) «What was the context of the times when a certain book was written» Cypher query statement match(p:book{«title»}) – (background) return p,background .

6) «Other related information about a book» Cypher query statement match(p:book{«title»}) – (related) return p,related.

7) «a book in what time to write» Cypher query statement match(p:book{«title»}) – (time) return p,time.

8) «Relationship between an author and a book or other authors» Cypher query statement match (p:person{name: «person's name»}) return p.

After the first few steps of entity recognition, relationship extraction and Cypher statement query, we can realize that from the Neo4j database query to the corresponding entity node, and get the keywords of the

final answer. This has basically completed the Q&A system Q&A requirements, but in order to make the user experience better, the final need to query the entity nodes for packaging. For example, for the user question «Who is the author of romance of the Three Kingdoms?» Through the entity recognition to identify the «Romance of the Three Kingdoms», «the author of» two entities, after the relationship classification to determine its relationship category for the first category «who is the author of a book», after the After Cypher statement query to get «match (p: Person {«book name»}) – (author) return p, author», you can get the author information. Thus, providing the user with an answer that can be more easily understood.

This system uses WeChat public number as the interactive interface to provide users with question and answer services. We need to connect the WeChat public number interface, the WeChat public number as a window for users to ask questions, in the call WeChat public number passive message reply interface and customer service service message interface to further complete the front and back-end interaction, the students can follow the public number to ask questions, so as to get the answer to their own questions (Li, 2022).

Conclusion

This paper focuses on the design and implementation of a knowledge graph-based teaching-assisted Q&A system, aiming at solving the problems encountered by students in the course of pre-study of classical Chinese masterpieces, improving the pre-study effect, and reducing the repetitive Q&A burden of teachers. In this paper, the knowledge system of classical Chinese masterpieces is constructed using knowledge mapping technology, which provides accurate and comprehensive knowledge support for the Q&A system. At the same time, combined with the characteristics of the classical Chinese masterpieces course, an intelligent Q&A system for the course is designed and implemented, which is able to automatically retrieve relevant information from the knowledge base and generate answers according to the questions inputted by the users, so as to provide personalized learning support for the students. The teaching-assisted Q&A system based on knowledge mapping constructed in this paper not only enriches the form of online culture dissemination, but also provides a powerful support for the teaching of the course of classical Chinese masterpieces, which helps to improve the students' preview effect and the teachers' teaching efficiency. In the future, we will continue to improve the system and explore more application scenarios to provide more convenient and efficient support for students' learning and teachers' teaching.

References

1. An T., Chen Y., Chen Y., Ma L., Wang, J., Zhao, J. A machine learning-based approach to ERα bioactivity and drug ADMET prediction // Front. Genet. 2023. № 13(108). pp. 72-73.
2. Anandika A., Chakravarty S., Paikaray B.K. Named entity recognition in Odia language: a rule-based approach // International journal of reasoning-based intelligent systems. 2023.
3. Cauteruccio F., Kou Y. Investigating the emotional experiences in eSports spectatorship: The case of League of Legends // Inf. Process. Manag. 2023. № 60. pp. 103-516.
4. Do Ph., Truong H.V. Phan. Developing a BERT based triple classification model using knowledge graph embedding for question answering system // Applied intelligence. 2021 (prepublish).
5. Feiran W., Bo T., Yue Ch., Xiyang F., Guowei J., Haiyan W., Gang Ch., Zhuangzhuang Sh. A visual knowledge map analysis of mine fire research based on CiteSpace // Environmental science and pollution research international. 2022. № 29(51).
6. Feng P., Ningzhang B. Entity alignment interaction model based on Chinese RoBERTa // Applied sciences. 2024. № 14(14).
7. Hou X., Zhu Ch., Li Yu., Wang P., Peng X. Question answering system based on military knowledge graph[P] // University of Electronic Science and Technology of China. 2022.
8. Li W. An improved knowledge graph question answering system for English teaching // Mobile information systems, 2022.
9. Li X., Raga R.C., Xuemei Shi. GloVe-CNN-BiLSTM model for sentiment analysis on text reviews // Journal of sensors. 2022.

10. Ping Ch., Jian Yi, Yue Chao Zh. Intelligent question answering system by deep convolutional neural network in finance and economics teaching // Computational intelligence and neuroscience. 2022.
11. Qian Sh., Ming Xu, Bin Q., Pengbin L., Junjian H. Intelligent question answering system based on machine reading comprehension // Journal of physics: conf. series. 2021. № 2050(1).
12. Rocha Rúben R., Carla M.S.E, Anderson C.R. Embeddedness in internationalization: Knowledge map and research agenda proposal // Journal of general management. 2022. № 48(1).
13. Zheng Zh., Liu M.,Weng Zh. A Chinese BERT-based dual-channel named entity recognition method for solid rocket engines // Journal electronics. 2023.

Разработка системы вопросов и ответов на основе графа знаний для оказания помощи в обучении

Пин Фэн
Аспирант
Чанчуньский университет
Чанчунь, Китай
fengping@ccu.edu.cn
ORCID 0000-0003-4865-1454

Сюй Сыци
Магистр
Чанчуньский университет
Чанчунь, Китай
2786265330@qq.com
ORCID 0000-0000-0000-0000

Инин Ван
Магистр
Чанчуньский университет
Чанчунь, Китай
2411207712@qq.com
ORCID 0000-0000-0000-0000

Поступила в редакцию 05.12.2024
Принята 26.01.2025
Опубликована 15.02.2025

УДК 004.421.2
DOI 10.25726/b2852-1767-1381-l
EDN DYIUOV
ВАК 5.8.2. Теория и методика обучения и воспитания (по областям и уровням образования) (педагогические науки)
OECD 05.03.HB. EDUCATION, SCIENTIFIC DISCIPLINES

Аннотация

Целью данной статьи является создание системы вопросов и ответов на основе графов знаний, которая ориентирована на удовлетворение потребностей студентов в предварительном просмотре курса, посвященного шедеврам классической китайской литературы. Объединяя отображение знаний с несколькими ресурсами, система способна обогатить форму традиционной культурной коммуникации и

предоставлять целенаправленные услуги по заданию вопросов и ответов. Задача исследования – помочь студентам подготовиться к курсу заранее, получить своевременные ответы на вопросы, снизить нагрузку на преподавателей, повысить эффективность преподавания в классе и качество обучения студентов, а также изучить пути распространения традиционной китайской культуры.

Ключевые слова

картирование знаний; система вопросов и ответов; классические китайские шедевры; распространение традиционной культуры.

Эта статья подготовлена при финансовой поддержке следующих проектов: Исследовательский проект Ассоциации высшего образования провинции Цзилинь на 2023 год (исследование по интеграции искусственного интеллекта и наследованию превосходной традиционной китайской культуры). Номер одобрения проекта: JGJX2023D397. Исследовательский проект провинции Цзилинь по реформированию профессионального образования и образования взрослых в 2023 году (исследование по глубокой интеграции технологий искусственного интеллекта во все процессы преподавания и управления) Номер утверждения проекта: 2023ZCY352.

Список литературы

1. An T., Chen Y., Chen Y., Ma L., Wang, J., Zhao, J. A machine learning-based approach to ERα bioactivity and drug ADMET prediction // Front. Genet. 2023. № 13(108). pp. 72-73.
2. Anandika A., Chakravarty S., Paikaray B.K. Named entity recognition in Odia language: a rule-based approach // International journal of reasoning-based intelligent systems. 2023.
3. Cauteruccio F., Kou Y. Investigating the emotional experiences in eSports spectatorship: The case of League of Legends // Inf. Process. Manag. 2023. № 60. pp. 103-516.
4. Do Ph., Truong H.V. Phan. Developing a BERT based triple classification model using knowledge graph embedding for question answering system // Applied intelligence. 2021 (prepublish).
5. Feiran W., Bo T., Yue Ch., Xiyang F., Guowei J., Haiyan W., Gang Ch., Zhuangzhuang Sh. A visual knowledge map analysis of mine fire research based on CiteSpace // Environmental science and pollution research international. 2022. № 29(51).
6. Feng P., Ningzhang B. Entity alignment interaction model based on Chinese RoBERTa // Applied sciences. 2024. № 14(14).
7. Hou X., Zhu Ch., Li Yu., Wang P., Peng X. Question answering system based on military knowledge graph[P] // University of Electronic Science and Technology of China. 2022.
8. Li W. An improved knowledge graph question answering system for English teaching // Mobile information systems, 2022.
9. Li X., Raga R.C., Xuemei Shi. GloVe-CNN-BiLSTM model for sentiment analysis on text reviews // Journal of sensors. 2022.
10. Ping Ch., Jian Yi, Yue Chao Zh. Intelligent question answering system by deep convolutional neural network in finance and economics teaching // Computational intelligence and neuroscience. 2022.
11. Qian Sh., Ming Xu, Bin Q., Pengbin L., Junjian H. Intelligent question answering system based on machine reading comprehension // Journal of physics: conf. series. 2021. № 2050(1).
12. Rocha Rúben R., Carla M.S.E, Anderson C.R. Embeddedness in internationalization: Knowledge map and research agenda proposal // Journal of general management. 2022. № 48(1).
13. Zheng Zh., Liu M.,Weng Zh. A Chinese BERT-based dual-channel named entity recognition method for solid rocket engines // Journal electronics. 2023.