Application and Effectiveness of Improving Retrieval Systems Based on User Understanding in Smart Archive Management Systems

Chao Yan
General Administration Office, YanCheng Polytechnic College, Yancheng, 224005, China

***Abstract*—In traditional archive management systems, keyword-based retrieval systems often fail to meet users' personalized and precise retrieval needs. To solve this problem, a knowledge graph is first constructed using bidirectional long short-term memory networks and conditional random fields and combined with user understanding-based semantic retrieval to obtain an improved personalized retrieval system. The research results show that the improved personalized retrieval system has significantly better retrieval accuracy and recall rate than traditional retrieval systems. The improved personalized retrieval system has retrieval accuracy rates of 90.24%, 89.65%, 87.52%, 96.33%, and 95.18% for students, civil servants, demobilized soldiers, law enforcement personnel, and retirees, respectively, and recall rates of 89.35%, 91.57%, 89.34%, 97.54%, and 96.63%, respectively. Applying it to the smart archive management system, the accuracy of archive retrieval, personalized recommendation accuracy, response time, and user satisfaction are significantly better than conventional management systems. The improvement and introduction of personalized retrieval systems based on user understanding and knowledge graphs have achieved significant results.**

***Keywords*—Knowledge graph; user understanding; retrieval system; smart archive management; BiLSTM-CRF**

I. Introduction

The rapid development of the information age has led to increasing attention to the Smart Archive Management System (SAMS). SAMS is a system that utilizes modern information technologies such as big data, cloud computing, the Internet of Things, artificial intelligence, etc. to intelligently manage archive information. It not only includes digital archival information input, storage, and retrieval, but also includes the protection, analysis, and utilization of archival digital resources, as well as online services and interactive functions. This system aims to improve the efficiency and quality of archive management while facilitating users to quickly and accurately obtain the required archive information. SAMS is mainly used by government agencies, enterprises, educational institutions, research institutions, archive workers, and the general public. It is responsible for managing public archives and providing public services, managing enterprise documents and commercial archives, managing academic and administrative archives of schools, and querying and utilizing archive resources. Traditional archive management methods often rely on paper documents and manual operations, which have problems such as low management efficiency and difficulties in information sharing. In the context of intelligence and automation, providing a more accurate, intelligent, and personalized retrieval system has become an important challenge for smart archive management systems. Traditional retrieval systems use keyword matching, which often makes it difficult to accurately understand the user's query intent and needs. Because relying solely on keyword matching, the system cannot understand the true intention behind the user's query, resulting in the returned results may not be accurate enough or do not meet the user's expectations. And traditional systems lack the ability for personalization. This method does not consider user behavior patterns and preferences, and cannot provide targeted search results, which cannot meet the personalized needs of different users. SAMS has replaced the old AMS because it provides a more efficient and comprehensive management approach. SAMS can improve efficiency, automate processing processes, reduce manual operations, and improve work efficiency. Meanwhile, digital archive information can be quickly retrieved, improving the speed of information acquisition. Cloud storage technology enables remote access and sharing of archival information, advanced data security technologies provide better protection of archival information, and data analysis tools can help decision-makers better understand archival information and make wiser decisions. In the context of intelligence and automation, providing a more accurate, intelligent, and personalized retrieval system has become an important challenge for SAMS [1-2]. Traditional retrieval systems use keyword matching, which often makes it difficult to accurately understand users' query intentions and needs. Therefore, providing personalized and accurate retrieval services has become particularly important for meeting the needs of users. User behavior mainly focuses on the operational behavior of users in the system. User preferences refer to the degree of user preference for the system or information, while semantic understanding is a deep understanding of user input or needs. Choosing to model user behavior is to improve the operational process and efficiency of the system. Choosing to model user preferences is for personalized recommendations and customized services. Choosing to model semantic understanding is to more accurately understand user needs and provide accurate retrieval results. The improvement of retrieval systems based on user understanding can provide users with more adaptive and personalized retrieval results by deeply understanding their query intentions and information needs, as well as modeling and analyzing their behavior patterns and preferences [3-5]. This method can be improved from multiple perspectives, including modeling user behavior, user preferences, and semantic understanding.

Knowledge graph is a structured knowledge representation method that describes the relationships between entities in the real world. Knowledge graph can provide richer and more accurate semantic information, providing strong support for personalized retrieval. By matching users' query intentions with entities and relationships in the knowledge graph, the accuracy and effectiveness of retrieval can be effectively improved [6-7]. The study utilizes bidirectional long short-term memory networks and conditional random fields to construct a knowledge graph, and combines it with user understanding-based semantic retrieval to obtain an improved personalized retrieval system. We hypothesize that combining user understanding with knowledge graph can better understand users' real needs and provide more accurate and personalized archive retrieval services. Applying the improved personalized retrieval system to SAMS will have more impact. The innovation of this study is to use Bidirectional Long Short Term Memory Network (BiLSTM) and Conditional Random Field (CRF) to construct a knowledge graph, namely BiLSTM-CRF. This system can better capture the semantic associations between queries and archives, thus achieving more intelligent and accurate retrieval results. Research has shown that the improved personalized retrieval system has significantly better retrieval accuracy and recall than traditional retrieval systems. Although significant progress has been made in many aspects of existing archive management systems and data processing frameworks, there are still several unresolved research gaps. Firstly, most existing systems lack flexibility and adaptability to meet different user needs. The current system is usually designed to be fixed and difficult to adjust according to the specific needs of different user groups, resulting in poor efficiency and effectiveness in handling cross-domain and diverse data. This limitation is particularly evident in application scenarios that require cross institutional collaboration and cross-platform integration. Secondly, there is a lack of in-depth research on the management of data bias and the handling of dynamic content changes. Although deviation detection has been preliminarily applied in some data analysis fields, its application in data archiving and long-term management is still insufficient. In addition, how to efficiently detect and adapt to changes in archived content to ensure the freshness and relevance of data remains an urgent challenge that needs to be addressed. This study aims to fill these research gaps and propose a flexible and adaptive archive management system framework that can not only dynamically adjust according to the needs of different user groups, but also effectively handle potential biases and content changes in data. By introducing advanced machine learning algorithms and dynamic update mechanisms, this study provides a new solution to improve the accuracy and fairness of data management. This will not only help improve the performance and practicality of existing systems but also promote the development of data archiving management and facilitate cross-industry and interdisciplinary data sharing and cooperation.

II. Related Work

Smart Archive Management (SAM) is a method that utilizes artificial intelligence and automation technology to improve the efficiency and quality of archive management. This technology can draw on the concepts and methods of knowledge graph and semantic retrieval to provide more intelligent and valuable archival services. Li Y proposed a cross domain recommendation method based on user preference perception and graph attention networks to improve the performance of knowledge graph recommendation systems. This method utilized a graph embedding model to obtain preference aware entity embedding, and combined preference features for personalized recommendation. It could effectively alleviate the problem of data sparsity and had good comprehensive performance, which can generate good personalized recommendation results [8]. Sui Y have constructed a model based on causal filters to improve the performance of knowledge graph question answering systems. This model utilized a data-driven approach to cause and effect interference in the relationship representation space, and disconnected other confounding factors in the knowledge embedding space through causal intervention. This method had good robustness and could effectively improve the comprehensive performance of the knowledge graph question answering system [9]. HX A et al. designed a knowledge graph model based on semantic fusion to address the issue of inaccurate semantic expression in knowledge graphs. It extracted subject entities by constructing a deep transformation model and generated candidate answer sets using dynamic candidate paths. This model could effectively solve the problem of inaccurate semantic expression in the knowledge graph, and improve the overall performance of the model [10]. Wang et al. designed a Top-k semantic aware query method based on semantic boundary perception to improve the accuracy of Knowledge Graph Star Query (KGSQ). This method utilized boundary deletion for matching with low semantic similarity, which can improve the accuracy of KGSQ [11].

User understanding can improve retrieval efficiency and accuracy, enhance user experience, and promote the development of personalized services. A deep understanding of user needs and behaviors can help improve the data analysis and processing capabilities behind retrieval systems, enabling them to provide more accurate and effective decision support, such as optimizing resource allocation and guiding policy formulation. In order to explore the impact of clarifying questions on user behavior and the ability to identify relevant information, Zou J utilized implicit interaction data and explicit user feedback to discover that high-quality clarifying questions can improve user performance and satisfaction [12].

Attention mechanisms and models such as BiLSTM can be applied to different stages of semantic retrieval tasks, such as semantic modeling correlation evaluation of queries and document sequences, to improve the accuracy and effectiveness of retrieval. Dong X et al. proposed a cross-modal graph attention method that combines recursive gated neural networks and attention mechanisms to address the semantic gap issue in data in cross-modal retrieval. This method could eliminate heterogeneous gaps between modalities, effectively solving the semantic gap problem of data in cross-modal retrieval [13]. To improve the performance of knowledge graph embedding, Dai has constructed a knowledge graph representation learning model based on generative adversarial networks, using Wasserstein distance instead of traditional divergence. Wasserstein distance could effectively solve the problem of small gradients on discrete data and improve the embedding performance of knowledge graphs [14]. Liu T and other scholars constructed a Chinese WeChat click bait dataset to detect the security of hyperlinks in online social media, and constructed a WeChat click bait detection method based on BiLSTM and multiple features, introducing attention mechanisms. This method could effectively detect bait links and had a high accuracy [15]. Hu Y et al. proposed a semantic behavior prediction method to improve the prediction accuracy of traffic participants' behavior in an autonomous vehicle. This method constructed a universal semantic representation method suitable for the driving environment and converted it into a spatiotemporal semantic map to infer the internal relationship between the two. This method had high prediction accuracy and was beneficial for solving problems under different traffic conditions [16].

In summary, many scholars have conducted research on knowledge graphs, semantic retrieval, and smart archive management, but how to combine them is an unresolved issue [17-18]. Knowledge graphs, semantic retrieval, and smart archive management are interrelated concepts and technologies, and have important application value in the fields of information management and services. Attention mechanisms, bidirectional long short-term memory networks, and semantic retrieval are interrelated and play a role in semantic modeling, correlation evaluation, and information retrieval, aiming to improve the performance and effectiveness of retrieval systems. In view of this, this study will use a bidirectional long short-term memory network to construct a knowledge graph, improve personalized retrieval systems based on user understanding and knowledge graphs, and further enhance SAMS [19-20].

III. Improvement of Retrieval Function in SAMS

To improve the performance and user experience of the retrieval system in SAMS, this study improves the personalized retrieval system based on user understanding and knowledge graph. The first step is to construct a knowledge graph and use the BiLSTM-CRF model to complete the process of entity recognition and relationship extraction in the archival field. Then, combining user understanding and knowledge graph, users' search requirements for keywords can be better understood and met from the traditional semantic retrieval level.

A. Construction of Knowledge Graph Based on BiLSTM-CRF

The knowledge graph network is a relatively structured entity semantic knowledge network used to analyze and express the objective and real world, as well as the relationships among various concepts, entities, etc. A knowledge event graph network is composed of a set of relationships between entities in different time, space, and states. This network forms a new semantic based knowledge network model by describing the interrelationships between concepts and entity concepts [21-22]. Fig. 1 shows the process of constructing a knowledge graph.

Fig. 1 shows the knowledge graph constructed in this study, In Fig. 1, the process of constructing a knowledge graph, it is necessary to first obtain data, transform the data into knowledge, then fuse the knowledge, and provide knowledge graph visualization services. Finally, apply based on knowledge graph. However, this study designs and models the ontology hierarchy, relationships, and attributes of archives. Among them, attribute structure refers to the organizational description of archive metadata and content, clarifying the characteristics and values of each archive item. By establishing a correct ontology hierarchy model, relevant knowledge can be extracted and analyzed, and clear semantic descriptions and relationship explanations can be provided for knowledge concepts within the domain. The data sources of archival knowledge include data extracted from existing business systems and other fields, and the data is divided into institutional, semi-structured, and non-manual structured.

From the knowledge framework of archive ontology language, it can be seen that entity recognition is required for the institutions, characters, files, time, location, etc. of archive ontology language. The language relationships of archive ontology mainly include synonyms, antonyms, context, subclasses, and other relationships [23]. For entity recognition, this study intends to select the BiLSTM-CRF model for extracting relational data and adopt a distributed relational data extraction technology based on remote data supervision to reduce the model's dependence on manually annotated data.

Next, construct the BiLSTM-CRF model. Recurrent Neural Network (RNN) is a type of neural network that performs temporal processing on sequence data. Time series data refers to the data collected at different time points, which reflects the situation or degree to which something, phenomenon, etc. occurs over time. Long Short-Term Memory (LSTM) networks are a new type of RNN that aims to overcome the problems of gradient loss and gradient explosion in long sequence learning. Compared with RNN, LSTM has better performance over longer time series [24]. LSTM is composed of three parts: input gate, forgetting gate, and output gate, achieving protection and control of information. Fig. 2 shows the LSTM's basic structure.

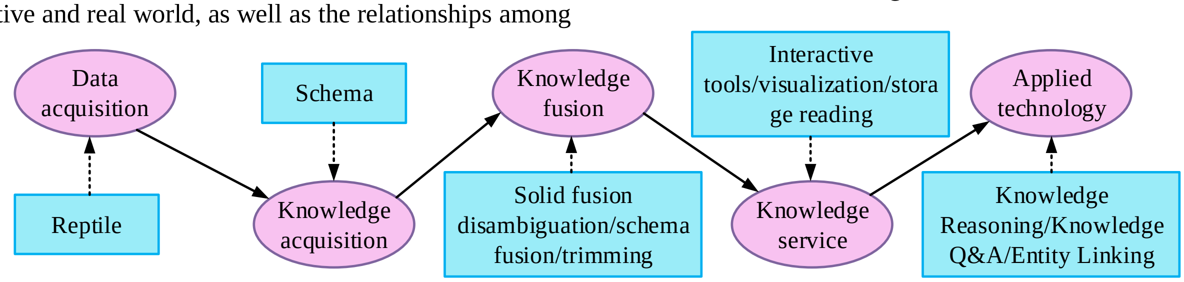

Fig. 1. The construction process of knowledge graph.

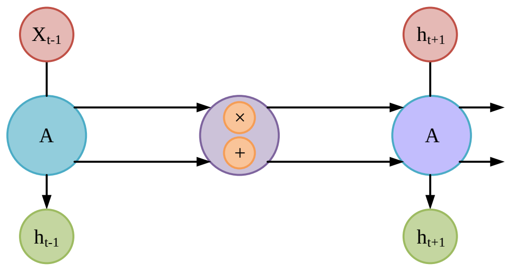

Fig. 2. The basic structure of LSTM.

BiLSTM is an RNN model that can process sequence data and capture long-term dependencies within the sequence. By running both a forward LSTM and a backward LSTM simultaneously, BiLSTM can utilize contextual information to better understand and encode input sequences. CRF refers to the conditional probability distribution modeling of another set of output variables given a set of input random variables. Its characteristic is to perform under the premise that the output variable satisfies the Markov random field [25].

The BiLSTM-CRF model is a neural network model used for sequence annotation tasks. It combines BiLSTM and CRF to simultaneously consider the dependency relationship between contextual information and labels. By utilizing the BiLSTM-CRF model, the process of entity recognition and relationship extraction in the archival field can be completed, as shown in Fig. 3.

After completing entity recognition and relationship extraction in the archival field, it is also necessary to store and retrieve the knowledge graph. Knowledge graphs are generally accessed through graph databases. After constructing and storing the enterprise archival knowledge graph in the graph database, it is necessary to combine advanced graph data retrieval technology to improve and enhance the efficiency of data query and processing in the enterprise archive knowledge graph. Its purpose is to provide support for achieving massive real-time dynamic information queries and data inference analysis.

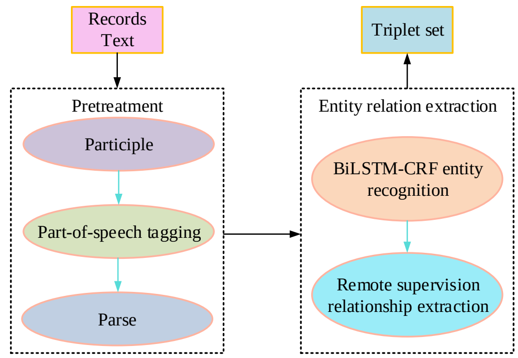

Fig. 3. Entity recognition and relationship extraction process in the field of archives.

B. *Improvement of Personalized Retrieval System Based on User Understanding and Knowledge Graph*

Traditional search engines mainly provide search results based on keyword matching, neglecting users' personalized needs and contextual information. However, users' search behavior and preferences are diverse, and traditional methods often find it difficult to accurately understand and meet these personalized needs. With the rapid growth of the internet scale and information content, it has become very important to better understand the semantic relationship between users and information [26]. Fig. 4 is a personalized retrieval system framework based on user understanding and knowledge graph.

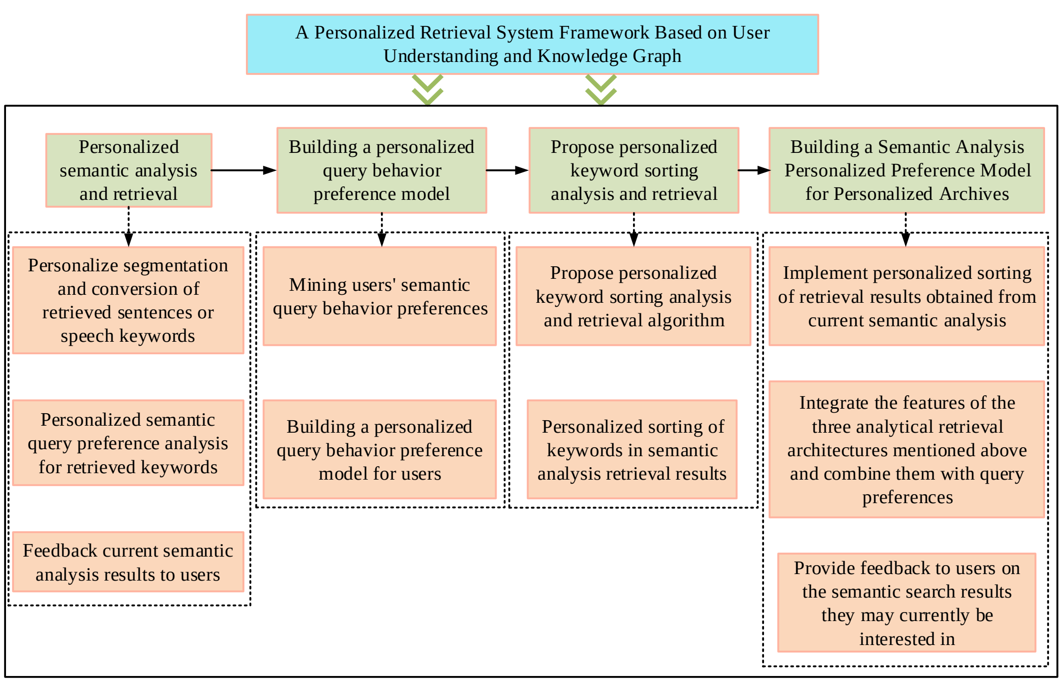

Fig. 4. A personalized retrieval system framework based on user understanding and knowledge graph.

From Fig. 4, it can be seen that in the personalized retrieval system framework, the first step is to conduct personalized semantic analysis retrieval. The second step is to build a personalized query behavior preference model. The third step is to propose a personalized keyword sorting analysis retrieval algorithm. The last step is to build a semantic analysis personalized preference model for personalized archives. Integrating the features of the three analytical retrieval architectures mentioned above, combining them with query preferences, to achieve personalized sorting of the retrieval results obtained from current semantic analysis, and providing feedback to users on the semantic retrieval results they may be interested in at present.

Traditional personnel file retrieval and query engines do have some limitations in providing personalized personnel file services. It is usually based on keyword matching and cannot fully understand users' personalized needs and contextual information [27]. Based on this, this study models personal information, enterprise information, and other information in the personnel information system based on user preference for personnel file query and data mining methods, achieving the retrieval of personal information in the personnel information system. Fig. 5 shows the mining framework operation process of user query preferences.

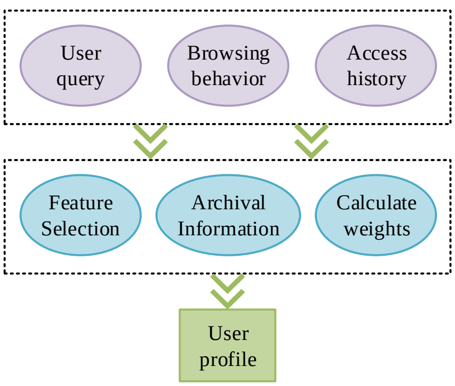

Fig. 5. The mining framework operation process of user query preferences.

In Fig. 5, the mining of user query preferences mainly includes seven steps, aiming to accurately analyze and represent user query preferences. These steps help to understand users' interests and preferences, and provide personalized search results, recommendations, and services based on these preferences.

The user query preference model is an algorithm model that describes the degree of interest of users in different types of archives. Each file object has a weight value that represents the user's level of interest in the file type. The higher the weight, the more interested the user is in this type of file [28]. This study uses probability models to compare and express user preferences for queries. It puts the file related data and file information related data from the user preference database into the user preference database, and generates a user preference probability distribution model for queries through analysis. w_i is defined as a weighting factor, which represents a user's preference for querying type i and file c_i in a short period of time. The calculation process is Eq. (1).

$$w_i = AF\left(c_i, \frac{x+1}{T}\right) \tag{1}$$

In Eq. (1), $AF\left(c_i, \frac{x+1}{T}\right)$ is used to describe the average value of file type c_i that users browse during period T. x represents the number of files browsed within T. By solving the weights of all file types in order, the user's query preference vector based on the file category can be obtained. The query preference characteristics of users are not fixed, and the required file types may also change under different job positions. So, after establishing a user's query preference model, there is a need for an algorithm that can update according to the user's query preferences. By setting an appropriate forgetting time, the problem of user query preference transfer can be solved. This study proposes a new user preference model based on forgetting factor using the Ebbinghaus memory curve. The expression is Eq. (2).

$$F(x) = \exp\left(-\frac{\log_2(t)}{f}\right) \tag{2}$$

In Eq. (2), t represents the time interval between the current date and the query preference creation date. f represents the half-life, which means that half of the time the user forgets will last for at least f days. Then, the user's query preference can be calculated based on the probability model and the user's query preference model [29].

Traditional keyword based retrieval methods cannot understand and meet users' retrieval needs for keywords at the traditional semantic level. Therefore, this study proposes a personalized retrieval method based on knowledge graph, which can better understand and meet users' search requirements for keywords from the traditional semantic retrieval level. Fig. 6 shows the semantic personalized retrieval process of archives based on knowledge graph.

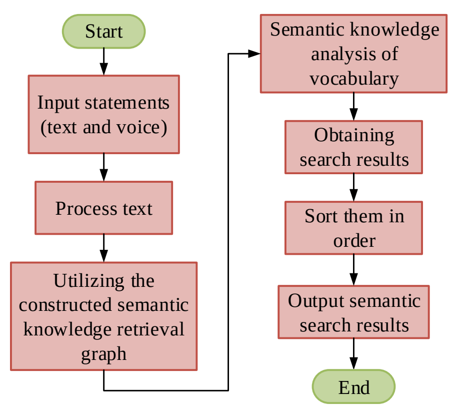

Fig. 6. The semantic personalized retrieval process of archives based on knowledge graph.

In Fig. 6, in the process of semantic personalized retrieval of archives based on knowledge graph, the first step is to input statements, including two types: text and speech. It is necessary to convert speech into text and then process the text. Afterwards, the constructed semantic knowledge retrieval graph can be used for semantic knowledge analysis of vocabulary. Then the search results are obtained and sorted to output semantic search results [30]. If the retrieval vocabulary is not included in the archive knowledge graph, semantic expansion and semantic mapping of the vocabulary are required, and the expanded retrieval vocabulary and the original retrieval vocabulary are added to synonymous concepts and segmentation specialized dictionaries respectively. If the current search term does not exist in the synonym dictionary, semantic similarity needs to be calculated, as shown in Eq. (3).

$$Sim(c,w) = \beta * Sim\_(c,w) + (1-\beta) * Sim\_2(c,w) \quad (3)$$

In Eq. (3), $Sim(c,w)$ is used to describe the minimum editing time distance between two words, calculated as shown in Eq. (4).

$$Sim_1(c,w) = 1 - \frac{ED(c,w)}{MLen(c,w)} \quad (4)$$

In Eq. (4), $Sim_2(c,w)$ is used to describe the distance from the editor, meaning the maximum semantic cosine length between and $\cos(e_c, e_w)$ represents the maximum semantic cosine distance between two words. Its similarity is equal to the distance between two synonyms and the embedded cosine similarity, as shown in Eq. (5).

$$Sim_2(c,w) = \cos(e_c, e_w) \quad (5)$$

In Eq. (5), e_c and e_w are embedded representations of words c and w. Then, by generating and constructing the inference graph of archival ontology knowledge, an archival knowledge inference system was constructed [31]. Through the Jena inference machine in the archive, semantic inference can be performed on the existing hierarchical knowledge relationships in the document. This enables semantic inference of the existing upper and lower knowledge relationships in archives, as well as better analysis and retrieval of the upper and lower semantics of documents. A key challenge faced by the system in the process of data archiving and management is how to handle potential biases in user data and changes in archive content. To effectively address these issues, the system has introduced a bias detection algorithm that can identify potential biases in the data. For example, in user-generated content, the system can identify anomalies or biases by analyzing data distribution and usage patterns. Through machine learning models, the system can automatically label these potential biased data and make corrections as necessary. This not only helps maintain the objectivity of data, but also improves the accuracy of data analysis results. The study selected 100 different types of individuals aged 18-45 for the experiment, all of whom were from the local archives bureau. The selected individuals were all from different professions and had significant occupational differences. The participants were divided into five groups, with 20 people in each group. Information was collected through a questionnaire survey. In the experiment data collection of this study, the network data are all from the Internet, and the files are stored separately in the database and web pages; the file data is provided by the local archives bureau, which scans paper files, converts them into PDF format electronic files, and saves them in a MySQL database.

IV. Application and Effect Analysis of Improving Retrieval System on SAMS

To verify the performance of the improved retrieval system, the performance of the knowledge graph based on BiLSTM-CRF and the retrieval performance of the preference semantic retrieval model based on user understanding are first analyzed, and the improved retrieval system was applied to the smart archive management system.

A. Performance Analysis of Improved Personalized Retrieval System Based on User Understanding and Knowledge Graph

In order to verify the performance of personalized retrieval systems based on user understanding and knowledge graph improvement, knowledge graph, user understanding, and improved personalized retrieval systems were analyzed in sequence. In the experimental data collection of this study, network data was sourced from the Internet, and archives were stored separately in databases and web pages. The piece of data is provided by the local archives bureau, which scans paper files, converts them into electronic files in PDF format, and saves them in a MySQL database. The file data mainly includes retrieval information for specific occupational groups. Table I shows the experimental system settings.

The experimental setup introduced in Table I can meet the experimental analysis of user understanding, knowledge graph, and retrieval system.

For the improvement of user understanding of the smart archive management system, specific occupational groups such as students, civil servants, demobilized soldiers, law enforcement personnel, and retirees were selected for application and effect analysis. These occupational groups represent users in society with different information needs and backgrounds in using archive management systems. Students need academic resources such as historical data and scientific research data, civil servants and law enforcement personnel need policy documents, laws and regulations, work files, etc. Demobilized soldiers and retirees need relevant welfare policies, personal service records, and other information. They need to obtain academic resources, research data, policy documents, welfare policies, laws and regulations, and other information from archives. Obtain information directly from users through methods such as questionnaire surveys, face-to-face interviews, and online interviews. Analyze the logs and usage records of the smart archive management system, and understand the user's usage habits and needs. Study specific cases of user groups using the system, including successful cases and problematic cases. Design representative tasks or scenarios for users from different professions to perform in a smart archive management system, such as retrieving specific information, using a certain service, etc. Collect data through system logs, user feedback, observation records, and other methods, including user operation behavior, task completion time, user satisfaction, etc. Using methods such as statistical analysis, content analysis, and behavioral analysis, analyze data from both quantitative and qualitative perspectives, identify user needs, evaluate system effectiveness, and identify system shortcomings. Firstly, to analyze the performance of the BiLSTM-CRF based knowledge graph and compare it with conventional knowledge graphs. The results are shown in Fig. 7.

Fig. 7(a) and Fig. 7(b) respectively represent the retrieval accuracy and recall of BiLSTM-CRF based knowledge graphs and conventional knowledge graphs. The former has retrieval accuracy rates of 92.34%, 88.61%, 90.22%, 90.08%, and 89.34% for students, civil servants, demobilized soldiers, law enforcement personnel, and retirees, respectively. The latter are 63.15%, 71.54%, 55.63%, 72.06%, and 72.65%, respectively. The recall rates of the former are 84.37%, 81.26%, 85.17%, 84.96%, and 82.33%, respectively, while the latter is 57.29%, 63.84%, 42.65%, 64.13%, and 70.19%, respectively. The retrieval accuracy and recall of knowledge graphs based on BiLSTM-CRF are significantly superior to conventional knowledge graphs. Next, analyzing the retrieval performance of the preference semantic retrieval model based on user understanding, and comparing it with conventional semantic retrieval models. The results are shown in Fig. 8.

Fig. 8(a) and Fig. 8(b) represent the retrieval accuracy and recall rates of the preference semantic retrieval model based on user understanding and the conventional semantic retrieval model, respectively. The former has retrieval accuracy rates of 93.64%, 90.12%, 88.47%, 95.24%, and 89.68% for students, civil servants, demobilized soldiers, law enforcement personnel, and retirees, respectively. The latter are 71.23%, 70.36%, 69.53%, 68.18%, and 71.62%, respectively. The recall rates of the former are 86.55%, 91.62%, 82.18%, 84.23%, and 96.35%, respectively, while the latter are 68.37%, 71.34%, 64.38%, 71.09%, and 80.24%, respectively. The experimental results indicate that, the retrieval accuracy and recall of the preference semantic retrieval model based on user understanding are significantly superior to conventional semantic retrieval models. Then to analyze the personalized retrieval system based on user understanding and knowledge graph improvement, compare it with traditional retrieval systems, and the results are shown in Fig. 9.

In Fig. 9, compared to traditional retrieval systems, the improved personalized retrieval system has better accuracy and recall. The improved personalized retrieval system has retrieval accuracy rates of 90.24%, 89.65%, 87.52%, 96.33%, and 95.18% for students, civil servants, demobilized soldiers, law enforcement personnel, and retirees, respectively. The retrieval accuracy of traditional retrieval systems is 79.63%, 78.27%, 74.57%, 80.24%, and 82.15%, respectively. The recall rates of the former are 89.35%, 91.57%, 89.34%, 97.54%, and 96.63%, respectively, while the latter is 72.14%, 74.54%, 70.68%, 79.27%, and 81.08%, respectively. The experimental results indicate that this study can improve the efficiency and user experience of smart archive management systems through user understanding-based retrieval systems, to achieve more accurate and personalized information retrieval and management.

TABLE I. SETTING OF EXPERIMENTAL SYSTEM ENVIRONMENT

Number	Project	Type	Name
(1)	Hardware requirements	Server	Hewlett Packard Enterprise
		Cloud platform	Microsoft Azure
(2)	Software requirements	Operating system	Windows
		Database Management System	MySQL
		Development environment	Python
(3)	Knowledge Graph Construction	Build Tools	OpenKG, Protégé, etc
(4)	Algorithms and Technologies	Algorithm	BiLSTM-CRF

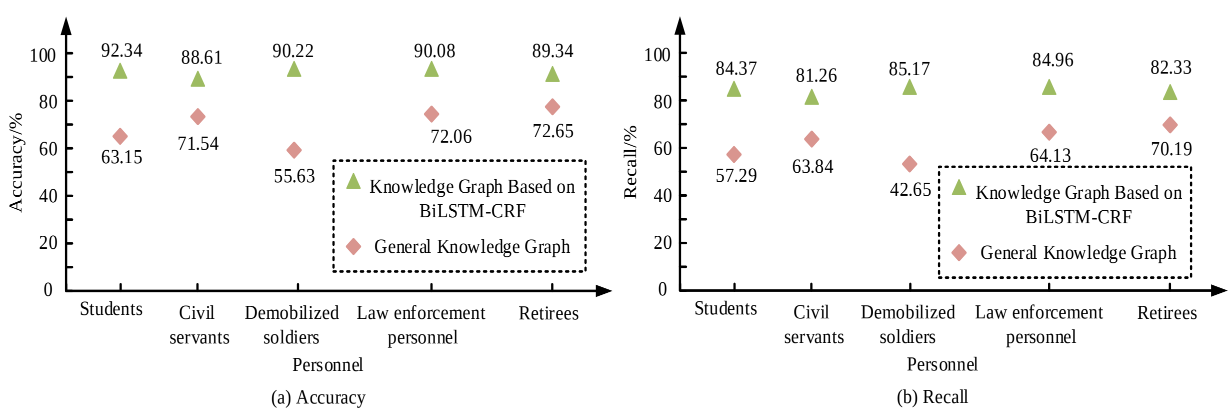

Fig. 7. Performance analysis of knowledge graph based on BiLSTM-CRF.

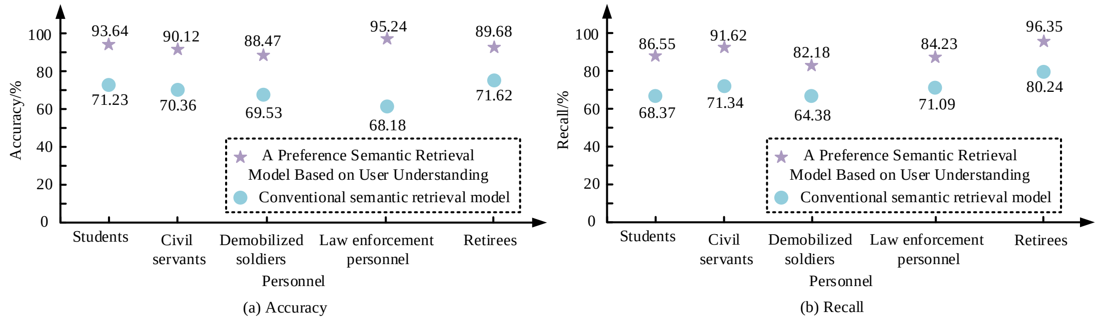

Fig. 8. Analysis of retrieval performance of preference semantic retrieval model based on user understanding.

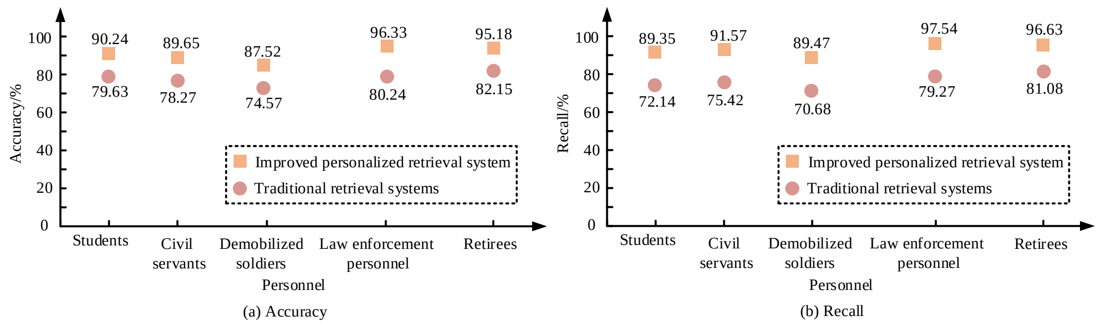

Fig. 9. Comparison of personalized vs. traditional retrieval systems, showing improvements in user understanding and knowledge graphs.

B. *Analysis of the Application Effect of Improved Retrieval System in SAMS*

In order to comprehensively evaluate the application effect of improving the retrieval system in the smart archive management system, first set goals, determine the user group participating in the evaluation, design representative retrieval tasks, and set a series of quantitative indicators to evaluate system performance. It is necessary to establish a control group in order to more accurately evaluate the improvement effect. In order to ensure the reliability of the experimental results, it is necessary to conduct experimental tests in a controlled environment and minimize the interference of external factors.

To select an archive dataset with a certain scale and diversity, and pre-process it into a format suitable for simulation analysis. 100 individuals aged 18 to 45 of different types were randomly selected to conduct experiments to analyze the application effect of improved retrieval systems in SAMS. Firstly, the accuracy of file retrieval and related personalized recommendations of the participants in the experiment were compared. The participants were divided into five groups of 20 people each, and the results are shown in Fig. 10.

In Fig. 10, the accuracy of file retrieval and related personalized recommendations in SAMS are significantly superior to conventional management systems. The accuracy of the former is more stable and smoother, while the accuracy of the latter shows a decrease. The experimental results indicate that the proposed system model has better performance. Next, to analyze the response time and user satisfaction of different personnel towards improving the retrieval system, as shown in Fig. 11.

In Fig. 11, the response time of SAMS is significantly shorter than that of conventional management systems, while user satisfaction is significantly higher than that of conventional management systems, indicating that the former has better overall performance. The response time of SAMS and conventional management systems is up to 8 seconds and 32 seconds respectively, with the highest user satisfaction of 92.34% and 71.42%, respectively.

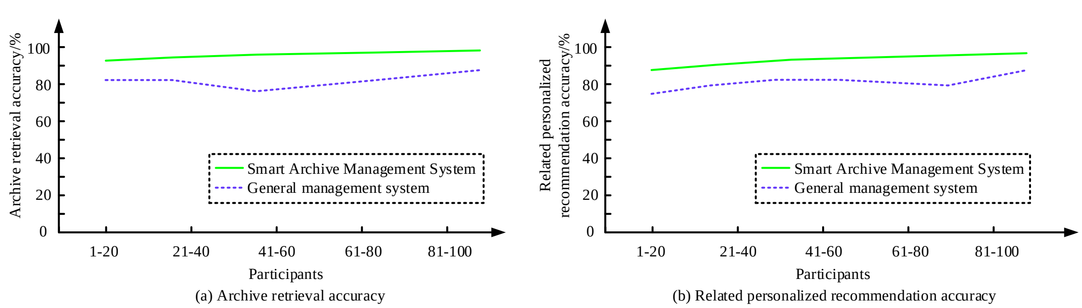

Fig. 10. Accuracy of archive retrieval and personalized recommendation.

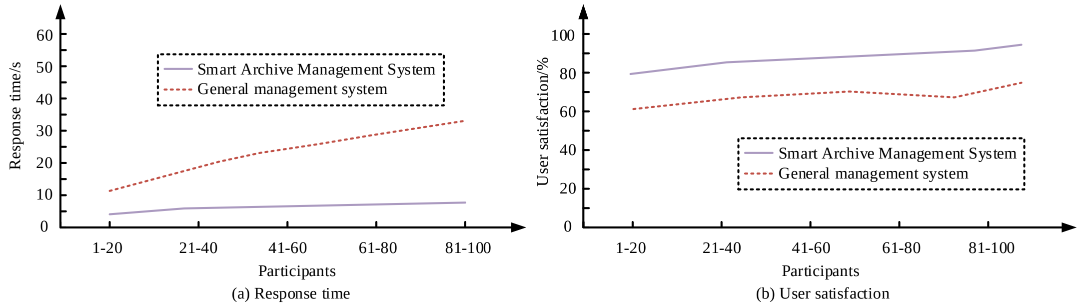

Fig. 11. Response time and user satisfaction of different personnel towards improving the retrieval system.

V. Discussions

In order to improve the performance and user experience of retrieval systems, the improvement of personalized retrieval systems based on user understanding and knowledge graphs has aroused the interest of researchers and practitioners. Research combines knowledge graphs and user understanding to construct personalized retrieval systems that meet user preferences. In order to meet the personalized needs of different users, the system continuously analyzes users' behavior patterns and preferences through machine learning algorithms and user feedback mechanisms. The system can provide personalized recommendations based on the user's operational history. In order to better understand the application of the system in real-world scenarios, research considers archive management environments of different scales and types. In large multinational corporations, this system can be used to manage millions of employee files and financial records. In government agencies, the system can process historical documents spanning decades. The scalability of the system enables it to handle the transition from paper documents to electronic documents and supports complex permission management to ensure the security of sensitive data. Through distributed storage and processing technology, the system is able to effectively manage and retrieve large amounts of documents, adapting to the constantly growing amount of data. In academic institutions, the archiving and management of research data require extremely high flexibility. This system meets the diverse data management needs of researchers by supporting various data formats and custom metadata tags.

In the actual implementation process, the system will face a series of technical challenges. For example, the real-time requirements of data processing may lead to increased system performance pressure, especially in high concurrency environments. In addition, data privacy and security issues are also a key challenge, and the system needs to effectively prevent potential data leakage risks. There are differences in technology acceptance among different user groups, which may affect the promotion and use of the system. With the continuous development of big data and artificial intelligence technology, the system should introduce more intelligent deviation detection and correction mechanisms to ensure the integrity and accuracy of data. Neglecting this aspect may lead to inaccurate data analysis results and even affect the effectiveness of decision-making. Therefore, I believe that future research should place greater emphasis on identifying and correcting data biases, and developing more robust data management strategies. In addition, the discussion on system scalability and adaptability also made me realize that relying solely on existing technological means is not enough. Interdisciplinary collaboration is needed to combine knowledge and technology from different fields in order to design more efficient and intelligent archiving management systems. This is not only a technical challenge, but also a management and policy issue that requires joint efforts from all parties to promote the development of this field.

VI. Future Work Prospects

This study has achieved preliminary results in the design and development of a flexible and adaptable archive management system, but in order to cope with constantly changing user needs and technological challenges, the following key areas still need to be further explored and improved in future work.

Firstly, improving the scalability of the system is one of the key focuses of future work. With the rapid growth of data volume and user numbers, the system must be able to effectively expand to support large-scale data storage and management. Future work will focus on developing more efficient data storage and retrieval algorithms, while exploring distributed storage and computing technologies to enhance system processing power and response speed. Through these technological means, we hope to achieve seamless system expansion without sacrificing performance. Secondly, in order to better meet the needs of different user groups, we plan to optimize the user interface and interaction design. Different users have different operating habits and needs, and in-depth user research will help collect and analyze user feedback to improve the user experience of the system. By introducing designs that are more intuitive and in line with user psychological models, the aim is to improve the usability and user satisfaction of the system. Future design optimization will include reorganizing information structures, simplifying operational processes, and adding personalized settings options. Thirdly, enhancing the intelligence level of the system is another important direction for future work. Introducing artificial intelligence and machine learning technologies to enable the system to automatically learn and adapt to user behavior patterns, thereby providing personalized services.

In summary, future research will focus on improving system scalability, optimizing user experience, and enhancing intelligence, addressing data bias issues, promoting interdisciplinary collaboration, and conducting long-term evaluations. Through these efforts, we look forward to further enhancing the adaptability and practicality of the archive management system, providing new research directions for the academic community, and offering effective solutions for practical applications. These future jobs will not only contribute to technological advancements, but also bring broader impacts to society.

VII. Conclusion

Traditional archive retrieval systems are often based on keyword matching and cannot understand users' query intentions and information needs. To improve the performance and user experience of retrieval systems, the improvement of personalized retrieval systems based on user understanding and knowledge graph has attracted the interest of researchers and practitioners. Providing personalized and accurate retrieval services has become particularly important for meeting the needs of users. This study combined knowledge graph with user understanding to construct a personalized retrieval system that meets user preferences. The experimental results showed that the retrieval accuracy and recall rate of knowledge graph based on BiLSTM-CRF were significantly superior to conventional knowledge graphs. The former had retrieval accuracy rates of 92.34%, 88.61%, 90.22%, 90.08%, and 89.34% for students, civil servants, demobilized soldiers, law enforcement personnel, and retirees, and recall rates of 84.37%, 81.26%, 85.17%, 84.96%, and 82.33%, respectively. The retrieval accuracy and recall rate of the preference semantic retrieval model based on user understanding were significantly superior to conventional semantic retrieval models. Compared with traditional retrieval systems, improving personalized retrieval systems had better accuracy and recall. The improved personalized retrieval system had retrieval accuracy rates of 90.24%, 89.65%, 87.52%, 96.33%, and 95.18% for students, civil servants, demobilized soldiers, law enforcement personnel, and retirees, respectively. Compared with conventional management systems, SAMS had more comprehensive performance, with a maximum response time of 8 seconds and 32 seconds, and the highest user satisfaction of 92.34% and 71.42%, respectively. This indicates that the improved personalized retrieval system has good performance and has good application effects in SAMS.

A SAMS can integrate information from multiple data sources and connect them in a graphical structure. This type of link helps to discover the relationships between data, providing a more comprehensive and comprehensive perspective. Through semantic technology, SAMSs can achieve intelligent search and question answering functions. Users can ask questions through natural language, and the system will understand the meaning of the questions and find relevant information in the graph network, thereby providing accurate answers. A SAMS can help recommendation systems better understand user needs and interests. By analyzing user behavior and preferences, combined with information from the knowledge graph, recommendation systems can provide more personalized and accurate recommendation results. A SAMS can help computers understand and analyze unstructured data such as text and images. By mapping this data to a knowledge graph network, the system can understand the meaning of the data and extract useful information from it.

There are still some shortcomings in this study, such as a small sample size. Subsequent studies will expand the sample size to verify the stability of research accuracy and other indicators.

References

[1] Xu H, Huang C, Wang D. Enhancing semantic image retrieval with limited labeled examples via deep learning. Knowledge-Based Systems, 2019, 163(JAN.1):252-266.

[2] Feiyan Zhang. Thinking on The Information Construction and Standardized Management of University Construction Engineering Archives. Acta Informatica Malaysia. 2023; 7(2): 105-107.

[3] Almousa M, Benlamri R, Khoury R. Exploiting non-taxonomic relations for measuring semantic similarity and relatedness in WordNet. Knowledge-based systems, 2021, 212(Jan.5):106565.1-106565.19.

[4] Saleh Ahmed Jalal Siam, Mubashshir Bin Mahbub. Investigating User Perceptions of Ai Technology and Its Ethical Implications on Employment Dynamics and Bias. Acta Informatica Malaysia. 2024; 8(1): 22-25.

[5] Girang Permata Gusti, Hilda. The Potential Use of Quick Response (Qr) Codes in Mobile Banking: An Analysis of Implementation and Its Impact on User Experience. Malaysian E Commerce Journal. 2023; 7 (1): 54-57.

[6] Huang W, Mao Y, Yang Z, L Zhu, J Long. Relation classification via knowledge graph enhanced transformer encoder. Knowledge-based systems, 2020, 206(Oct.28):106321.1-106321.10.

[7] Tuo Shi, Danyang Li, Qi Zhang. Research On the Construction of A Knowledge Graph Of Interaction Risk Between Home Invasion Theft Offenders And Victims Based On Information Extraction Technology. Acta Informatica Malaysia. 2024; 8(1): 01-04.

[8] Li Y, Hou L, Li J. Preference-aware Graph Attention Networks for Cross-Domain Recommendations with Collaborative Knowledge Graph. ACM transactions on information systems, 2023, 41(3):80.1-80.26.

[9] Sui Y, Feng S, Zhang H, J Cao, L Hu, N Zhu. Causality-aware Enhanced Model for Multi-hop Question Answering over Knowledge Graphs. Knowledge-based systems, 2022, 250(Aug.17):1-16.

[10] HX A, SW B, MT A, LW A, XLA C. Knowledge Graph Question Answering with semantic oriented fusion model. Knowledge-Based Systems, 2021, 221(Jun.7):106954.1-106954.10.

[11] Wang Y, Xu X, Hong Q, J Jin, T Wu. Top- k star queries on knowledge graphs through semantic-aware bounding match scores. Knowledge-Based Systems, 2020, 213(2):106655.1-106655.13.

[12] Zou J, et al. (2023), "Users Meet Clarifying Questions: Toward a Better Understanding of User Interactions for Search Clarification", ACM transactions on information systems, Vol. 41 No. 1, pp.1.1-1.25.

[13] Dong X, Zhang H, Dong X, X Lu. Iterative graph attention memory network for cross-modal retrieval. Knowledge-Based Systems, 2021, 226(6):107138.1-107138.12.

[14] Dai Y, Wang S, Chen X, C Xu, W Guo. Generative adversarial networks based on Wasserstein distance for knowledge graph embeddings. Knowledge-Based Systems, 2019, 190(2):105165.1-105165.12.

[15] Liu T, Yu K, Wang L, X Zhang, H Zhou, X Wu. Clickbait detection on WeChat: A deep model integrating semantic and syntactic information. Knowledge-based systems, 2022, 245(Jun.7):108605.1-108605.11.

[16] Hu Y, Zhan W, Tomizuka M. Scenario-Transferable Semantic Graph Reasoning for Interaction-Aware Probabilistic Prediction. IEEE transactions on intelligent transportation systems, 2022, 23(12):23212-23230.

[17] Wozniak P A (2022). Hybrid Electric Vehicle Battery-Ultracapacitor Energy Management System Design and Optimization. Elektronika ir Elektrotechnika, Vol. 28 No. 1, pp.4-15.

[18] Kili S, Firat M (2023), zdemir.A Novel Current Condition Assessment System for Sustainable Management and Operation of Wastewater Treatment Plants.Journal of Environmental Engineering, 2023, Vol. 149 No. 6, pp.241-256.

[19] Du W, et al (2023). Sequential patent trading recommendation using knowledge-aware attentional bidirectional long short-term memory network (KBiLSTM). Journal of Information Science, Vol. 49 No. 3, pp.814-830.

[20] Jin H, Bao Z, Chang X, et al (2023). Semantic segmentation of remote sensing images based on dilated convolution and spatial-channel attention mechanism. Journal of Applied Remote Sensing. Vol. 17 No. 1, pp.016518-1- 016518-17.

[21] Efrén Rama-Maneiro, Vidal J C, Lama M. Collective disambiguation in entity linking based on topic coherence in semantic graphs. Knowledge-Based Systems, 2020, 199(Jul.8):105967.1-105967.14.

[22] Nie Z, Zheng S, Liu Y, Z Chen, S Li, K Lei, F Pan. Automating Materials Exploration with a Semantic Knowledge Graph for Li-Ion Battery Cathodes. Advanced functional materials, 2022, 32(26):2201437.1-2201437.7.

[23] Ayetiran E F, Sojka P, Novotn, Vít. EDS-MEMBED: Multi-sense embeddings based on enhanced distributional semantic structures via a graph walk over word senses. Knowledge-Based Systems, 2021, 219(May.11):106902.1-106902.14.

[24] Fares M, Moufarrej A, Jreij E, J Tekli, W Grosky. Unsupervised word-level affect analysis and propagation in a lexical knowledge graph. Knowledge-Based Systems, 2019, 165(FEB.1):432-459.

[25] Zhu Y, Hu L, Ning N, W Zhang, B Wu. A lexical psycholinguistic knowledge-guided graph neural network for interpretable personality detection. Knowledge-based systems, 2022, 249(Aug.5):108952.1-108952.14.

[26] K.-J. C, Liu Z, Lu H, J Zhang. Heterogeneous graph convolutional network with local influence. Knowledge-based systems, 2022, 236(Jan.25):107699.1-107699.10.

[27] Chen Z, Zhao X, Liao J, X Li, E Kanoulas. Temporal knowledge graph question answering via subgraph reasoning. Knowledge-based systems, 2022, 251(Sep.5):109134.1-109134.11.

[28] Rao V, Dai P, Singla S. Structural fragmentation in scene graphs. Knowledge-Based Systems, 2020, 211(Jan.9):106504.1-106504.11.

[29] Fang Y, Ren Y, Park J H. Semantic-enhanced discrete matrix factorization hashing for heterogeneous modal matching. Knowledge- Based Systems, 2020, 192(Mar.15):105381.1-105381.13.

[30] Fang Y, Luo B, Zhao T, He D, Jiang B, Liu Q. ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting. CAAI Transactions on Intelligence Technology, 2022, 7(4): 744-757.

[31] Guo Y, Mustafaoglu Z, & Koundal D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. Journal of Computational and Cognitive Engineering, 2022, 2(1), 5–9.