Construction and Application of a Job-Position Knowledge Graph in the Library and Information Field Based on a Competency Model*

Shi Yali¹ Liang Miao² Xiong Xinni¹ Zhang Lixia¹

2026-07-24

ChinaRxiv

Abstract

To address problems such as information blockage and asymmetry in job information retrieval in the field of library and information science, this paper attempts to construct a job knowledge graph for the field based on a competency model, and completes the architectural design and functional development of its application platform. First, online recruitment information in the field of library and information science is used as raw data, and tools such as Python are employed to conduct cluster analysis of job information. Job competencies in the field are divided into four dimensions: professional knowledge, job skills, personality traits, and professional literacy, covering 46 job competency indicators, including decision-making and analysis ability, data analysis ability, and language expression ability. Second, an ontology is created with Protégé, and a job knowledge graph for the field of library and information science is constructed using text mining methods. Finally, an application platform for the job knowledge graph in the field of library and information science is designed and subjected to scenario-based functional testing. The test results show that the application platform can effectively improve job seekers' information retrieval efficiency and has decision-making reference value for optimizing job recommendation services on recruitment platforms and promoting innovation in talent training models.

Full Text

[**Professional Practice · Management**]

Construction and Application of a Job-Position Knowledge Graph in the Library and Information Field Based on a Competency Model*

Shi Yali1 Liang Miao2 Xiong Xinni1 Zhang Lixia1
1. School of History and Culture, Hubei University, Wuhan, 430062
2. Huanggang Vocational and Technical University, Huanggang, 438000

Abstract To address problems such as information blockage and information asymmetry in the retrieval of job information in the library and information field, this paper attempts to construct a job-position knowledge graph for the library and information field based on a competency model, and completes the architectural design and functional development of its application platform. First, online recruitment information in the library and information field is used as the raw data, and tools such as Python are employed to conduct cluster analysis of job information. Job competencies in the library and information field are divided into four dimensions—professional knowledge, job skills, personality traits, and occupational qualities—covering 46 competency indicators, including decision-analysis ability, data-analysis ability, and language-expression ability. Second,

an ontology is created with Protégé, and a job-position knowledge graph for the library and information field is constructed using text-mining methods. Finally, an application platform for the job-position knowledge graph in the library and information field is designed, and scenario-based functional tests are carried out. The test results show that the application platform can effectively improve job seekers' information retrieval efficiency and has decision-making reference value for optimizing job recommendation services on recruitment platforms and promoting innovation in talent-cultivation models.

[**Keywords**] positions in the library and information field; competency model; knowledge graph; text mining; application platform

[**CLC Classification Number**] G250 [**Document Code**] A [**Article Number**] 1003-7845(2026)03-0034-10

[**Citation Format**] Shi Yali, Liang Miao, Xiong Xinni, et al. Construction and Application of a Job-Position Knowledge Graph in the Library and Information Field Based on a Competency Model [J]. *Work of University Libraries*, 2026(3): 34-43.

Introduction

The *Outline of the Fifteenth Five-Year Plan for National Economic and Social Development of the People' s Republic of China*[1] clearly states that efforts should be made to enhance the operational capacity of vocational schools, build higher vocational institutions with distinctive characteristics, thoroughly implement the strategy of educational digitalization, and optimize public services for lifelong learning. Together, these deployments constitute the current policy framework for the reform and development of vocational education, providing a clear direction for improving the quality of vocational education and strengthening its capacity to serve economic and social development. The *Outline of the Plan for Building China into a Leading Country in Education (2024-2035)*[2] points out that the high-quality development of vocational education should be promoted in an all-round way, highly skilled talent should be cultivated, and an education mechanism integrating moral cultivation with technical training and combining work with study should be improved, so that talent cultivation closely meets the needs of the front line of industry. At present, the library and information field is facing the challenges of transformation in information services and changes in the disciplinary knowledge system[3]. Practitioners' career-development environment is also undergoing the reshaping of the digital-intelligent era, and job demands in the library and information field are becoming increasingly diversified. However, there is currently a disconnect between the innovative transformation of disciplinary education and the development of occupational needs. Graduates and practitioners in the library and information field also have an "information gap" in their understanding of job competencies, making it difficult for them to formulate scientific and effective career plans. Large-scale knowledge graphs oriented toward Internet search have been widely

applied in data analysis, intelligent search and recommendation, natural human-computer interaction, decision support, and other areas[4]. Applying knowledge-graph technology to the field of job retrieval and recommendation can, on the one hand, obtain job retrieval results centered on semantic relations, rapidly realize the screening and matching of job information, and improve the efficiency of job-information retrieval; on the other hand, through the visual presentation of semantic relations, it can provide an intuitive understanding of the professional knowledge, skills, and comprehensive qualities required of practitioners in the library and information field, thereby facilitating talent cultivation and the improvement of occupational literacy.

Based on this, the present study proposes applying knowledge-graph technology to job retrieval and recommendation in the library and information field, and conducting cluster analysis and visual presentation of competency indicators for relevant positions. This can not only help job seekers understand the current state of career development and talent demand within the field and indicate directions for their self-improvement, but also provide reference ideas for talent-cultivation institutions to optimize training programs and carry out professional skills training. At the same time, the job-competency model for the library and information field constructed in this study refines and further validates the application scenarios of the classic iceberg model and onion model. Compared with traditional keyword retrieval or vector representation, the knowledge-graph-based method of representing jobs shows stronger contextual association and explanatory capacity in semantic understanding and relational reasoning, thereby offering a new theoretical perspective for research on related topics.

---

* This article is one of the research outputs of the National Social Science Fund Youth Project "Research on Risk Identification and Prevention of Researchers' Data Cognition in the Context of AI-Generated Content" (Project No.: 24CTQ056).

Author profiles: Shi Yali, Ph.D., associate professor, doctoral supervisor; research interests include document and information organization, standardization of scientific data citation, and natural language processing. Liang Miao, master' s degree, assistant librarian; research interests include information organization and knowledge management. Xiong Xinni, master' s student; research interests include information management and information systems, and semantic processing. Zhang Lixia, master' s student; research interests include information management and information systems, and semantic processing.

Date received: 2025-12-11 **Responsible editor:** Jia Nan

1 Related Research

At present, related research is mainly concentrated on knowledge-graph semantic reasoning, conceptual-relation processing, semantic processing, knowledge fusion, conversational retrieval, and other areas. Specifically, in semantic reasoning and conceptual-relation modeling, existing studies focus on refinement methods for knowledge graphs and supporting evaluation mechanisms, aiming to improve the application effectiveness of knowledge graphs in complex queries and logical reasoning and to enable the effective identification of erroneous information [5]. Regarding knowledge fusion in multilingual and cross-cultural contexts, the positive role of knowledge graphs in cross-language plagiarism detection has been confirmed; on this basis, Abu-Salih B et al. [6] proposed a vector-form-based weighting scheme for conceptual relations. In semantic processing and knowledge fusion, to address intelligent decision-making problems in complex business environments and to clarify the key role of knowledge graphs in improving the effectiveness of decision-making systems, Yu T [7] constructed an intelligent decision-making framework integrating design science research theory with business knowledge graphs. In the fields of digital humanities and smart tourism, knowledge graphs can realize the structuring and semantic processing of massive unstructured historical materials, such as the construction of chronological-biography knowledge graphs [8] and the construction of tourism-attraction knowledge graphs based on multi-source data fusion [9]. In terms of knowledge graphs empowering conversational learning, the fields of education and interdisciplinary discipline development have applied knowledge graphs to collaborative learning analytics. By constructing conversational knowledge graphs, they visually present the paths of group knowledge construction and the role characteristics of different types of learners, thereby providing a scientific basis for refined instructional intervention [10]. In the interdisciplinary field of area studies urgently needed for national strategy, Zheng Chunrong et al. [11] constructed a domain knowledge graph for the discipline, realizing closed-loop management from knowledge generation to practical application, and accordingly proposed targeted teaching strategies. In the application of position knowledge graphs, He Chunhui et al. [12] proposed a ranking method based on BERT and cosine semantic similarity to improve the accuracy of job-position matching. Based on the competency model, Zhang Xinxin [13] conducted text mining on online recruitment information and constructed a knowledge graph of data-related positions, providing theoretical grounding and technical reference for knowledge-graph research in the library and information field.

In summary, existing research has achieved certain results in the theoretical foundations and technical applications of knowledge graphs, and explorations in digital humanities, smart tourism, discipline development, job-position matching, and other fields can provide literature references for this study. However, current research still has the following shortcomings. First, existing studies mostly focus on core areas such as semantic reasoning, dynamic updating, and cross-language integration, aiming to improve the application capabilities of

knowledge graphs in complex queries, logical reasoning, and multilingual knowledge fusion. Although these technologies have made significant progress in universality and frontier development, their in-depth implementation and large-scale application in specific vertical domains remain relatively limited. Second, although the application of knowledge graphs in the job-seeking field has made some progress, it lacks a unified data foundation, and application scenarios are relatively scattered. These shortcomings provide an entry point for this study. Taking online recruitment information in the library and information field as the raw data, this study constructs a position knowledge graph and conducts modeling research on this basis, followed by system development and application, with a view to providing new ideas for related theoretical research and domain development.

2 Research Design and Research Methods

2.1 Research Approach

This study is mainly divided into four stages: data acquisition and preprocessing. Online recruitment information in the library and information field is collected using the Octoparse crawler, and the obtained raw data are preprocessed.

Construction of a competency model for positions in the library and information field. Cluster analysis is performed on the preprocessed text data, and the feature words obtained through clustering are classified by dimension and subdivided into indicators in combination with the classic competency model.

Construction of a position knowledge graph in the library and information field. Based on the competency model, entity recognition, relation extraction, and knowledge fusion are performed on the preprocessed unstructured data. The generated relational triples are stored and queried in the Neo4j knowledge-graph database, and, on the basis of clarifying entity relations and attribute types, visual presentation is realized with the aid of the Protégé ontology-design software.

Design and functional testing of the knowledge-graph application platform. Based on user needs, the framework and functional modules of the application platform for the position knowledge graph in the library and information field are designed, including three major modules: position search, position recommendation, and self-directed learning; typical application scenarios are then selected for functional testing.

2.2 Research Methods

When constructing the position competency model, this study mainly uses the K-means cluster analysis method to extract feature words. The initial value of K for corpus partitioning is preset, and the optimal value of K is selected by combining the elbow method and the silhouette coefficient method. On this basis, the final clustering results are calculated, and the feature words are classified by dimension in combination with word meanings and the classic competency

model, thereby constructing the position competency model. In the entity-recognition stage, the ERNIE pre-trained model is used to identify entities in the corpus. This model integrates big data with multi-source knowledge and continuously acquires lexical, structural, and semantic knowledge from text data through continual learning technology, thereby optimizing model performance. It can learn the semantic representation of complete concepts, more comprehensively and accurately identify semantic information, and enhance the semantic expression effect of entities [14]. In the relation-extraction stage, the RoBERTa pre-trained model is used to extract entity relations. RoBERTa optimizes the BERT model by using a larger-scale training dataset and increasing the number and duration of training iterations, so that the data are trained more fully. At the same time, RoBERTa removes the next-sentence prediction task in BERT and adopts a dynamic masking mechanism for data training, thereby effectively improving the model' s pre-training performance [15].

2.3 Data Acquisition and Data Preprocessing

2.3.1 Data Acquisition

To ensure that the constructed position knowledge graph in the library and information field conforms to the actual situation of the employment market, this study, starting from the comprehensiveness and representativeness of the raw data, identifies the following data sources: position data from government agencies and public institutions, sourced from the National Civil Service and provincial civil service examinations

job-position tables. According to the *2023 Online Recruitment and Job-Seeking Industry Insight* released by Analysys[16] and the *2024 China Online Recruitment Industry Research Report* released by iResearch Consulting[17], 51job enjoys high brand recognition across multiple dimensions, including the number of positions, professional authority, influence, and reputation, and its overall satisfaction level leads that of other recruitment platforms. Therefore, this study selected recruitment data from 51job as the main source of enterprise data. To ensure data comprehensiveness, three categories of data from Gaoxiaorencai.net —"university recruitment," "government and public institutions," and "enterprise recruitment" —were additionally selected as auxiliary data sources.

The recruitment-platform data were crawled using the Octoparse collector. The main keywords for data collection were "library science," "archival science," "library and information," "information management," "information science," and so on, and retrieval formulas were constructed through combinations of search terms. It should be noted that, in the current recruitment market, the responsibilities of positions related to the library and information field have already gone beyond traditional disciplinary boundaries, showing the characteristics of occupational integration under "broad information management." From the perspective of the operability and completeness of data collection, this paper extends the "library and information field"to an occupational scope covering information-

management functions such as library management, archives management, and intelligence consulting, rather than limiting it to the disciplinary level. In the data-collection stage, keywords in recruitment information such as "position title," "address," "experience requirements," "educational background," "recruiting organization," "company nature," "size," "industry," "salary," and "competency requirements" were selected as collection fields, yielding a total of 5,963 records. Government data were collected manually: 2023–2024 national and regional examination position lists were downloaded from the "Hanshi Civil Service Examination Information Network" (all recruitment information was checked and corrected against the official recruitment announcements of the National Civil Service Administration and provincial human-resources and social-security authorities). In the "professional requirements" column, the same keywords as those used for online recruitment were adopted for data screening and collation, yielding a total of 668 records. Because the quality of the raw data varied, with problems such as duplicate position information, missing valid fields, and relatively low relevance of some positions to the library and information field, after data cleaning and deletion, a total of 5,295 valid records were retained.

2.3.2 Data Preprocessing

The preliminarily screened data show that the distribution of positions in the library and information field is relatively broad. To ensure the presentation effect of the subsequent graph, the data were classified with reference to the position categories in the *Occupational Classification Dictionary of the People' s Republic of China (2022 Edition)*[18]. Based on an interpretation of job responsibilities, the acquired data were divided into 20 categories of positions, including management researchers, information-systems analysis engineering technicians, and information-security engineering technicians.

According to the field content in the position information, the initially obtained data can be divided into structured data and unstructured data. Structured data generally refer to data with a certain logical and physical structure, stored in databases according to a specific format; unstructured data refer to data other than structured data, and are usually stored in textual form[19]. Information such as "position title," "address," "experience requirements," "educational background," and "recruiting organization" was treated as structured data for preprocessing. On the basis of ensuring the authenticity of the original data, missing fields were supplemented; for example, information for which the "salary" field was empty was uniformly filled in as "negotiable." The long-text fields in position information belong to unstructured data and mainly include two types of information: "job responsibilities" and "job requirements." During preprocessing, these long-text fields were first separated manually, and then word segmentation was performed. This study selected Jieba as the word-segmentation tool, and improved the segmentation effect by adding a custom dictionary, a synonym dictionary, and a stop-word dictionary. After segmentation, word frequencies were counted to preliminarily identify the core elements of job com-

petency in the library and information field, such as information-retrieval ability, decision-analysis ability, computer proficiency, and communication and coordination ability.

3 Construction of a Job-Competency Model for the Library and Information Field

3.1 Competency Theory and the Basis for Model Construction

The most representative competency model is the iceberg model, which is a professional evaluation model for workers' job competence and is also one of the classic models for distinguishing competent from incompetent workers[20]. Based on the iceberg model, Richard Boyatzis proposed the onion model, which divides competency from the inside outward into three layers: the innermost layer consists of solidified traits and motives; the middle layer consists of variable self-concept, social roles, and values; and the outermost layer consists of accumulable knowledge and skills, corresponding to the iceberg model' s change from implicit to explicit dimensions[21]. At present, these two models not only provide a clear framework for extracting competency indicators, but also systematically present the explicit and implicit dimensions of job-competency elements. They have been widely applied in analyses of job competency in fields such as medicine, engineering, financial analysis, and public administration. Some scholars have also carried out empirical studies on job competency in the field of information-resource management[22], such as studies of librarian competency in the process of library reading promotion[23] and the construction of competency models for cultural-and-creative librarians in China' s public libraries[24]. This study aims to mine job-competency indicators in the library and information field, and therefore falls within the typical scope of competency research. Existing studies have conducted empirical analyses from perspectives such as libraries, verifying the applicability of the two theoretical models in the library and information field and, to a certain extent, providing theoretical grounding and methodological reference for this study.

3.2 Word Vectorization

Before clustering analysis, the preprocessed text data must be vectorized; that is, textual symbolic information is transformed into numerical information in vector form. Clustering analysis on this basis can gather words with high semantic similarity, thereby achieving the purpose of extracting feature words. The Word2Vec word-embedding model is a model that can efficiently realize the conversion of text words into vectors. Its basic principle is that words with similar contexts should also have similar word vectors[25]. The Word2Vec model includes two implementation methods: the skip-gram model and the continuous bag-of-words model. The principle of the skip-gram model is to predict the words that may appear in the context through a given word; in this process, the center word is continuously updated. After this operation is repeated for

all texts, the word-vector trans—

results. Because in this process each word is predicted as a central word, computational complexity is increased; therefore, this model has the drawbacks of requiring many prediction iterations and high time cost. However, when the corpus contains a large number of low-frequency words, the results obtained by this model are more accurate. In contrast to the skip-gram model, the continuous bag-of-words model predicts a specific word by inputting words related to the context of that specific word, and has the advantages of high efficiency and fast speed. Taking into account the data volume and lexical characteristics of this study, this paper decides to use the continuous bag-of-words model as the word-vector model. In Python, the Word2Vec model is invoked to train the preprocessed unstructured data. After multiple rounds of testing, the window size is set to 2, the word-vector dimension is set to 200, the number of training iterations is set to 10, and the other parameters are kept unchanged. At this point, the model accuracy reaches the optimum, and word-vector results can be output.

3.3 Cluster Analysis

The vectorized dataset is read in Python, and the sklearn package is used for cluster analysis and to calculate the SSE value. In this study, the initial range of K is set to 4–9. The calculated SSE values show that the corresponding K value at the local elbow point is 6, where the decline in SSE begins to slow. On this basis, the silhouette coefficient is calculated to determine the optimal K value. In Python, the Silhouette function is used to calculate the silhouette coefficients for K values from 4 to 9, which are 0.471 014 26, 0.457 247 47, 0.373 845 3, 0.385 624 77, 0.372 457, and 0.361 360 73, respectively. The closer the silhouette coefficient is to 1, the better the clustering effect; when $K = 6$, the silhouette coefficient is relatively low. Taking the above two results into comprehensive consideration, 7 is selected as the optimal K value, and the clustering results for seven categories are then derived. According to the clustering results, the competency feature words to be extracted in this study are mainly concentrated in Categories 1, 2, 4, 5, and 7.

Based on the meanings of the feature words obtained from clustering, and in combination with iceberg model theory and onion model theory, the feature words are divided into four dimensions: professional knowledge, job skills, personality traits, and occupational literacy. The top 30 words by word frequency in each dimension are then counted separately. Through screening and integration, 46 competency indicators are finally extracted, thereby constructing the job competency model for the library and information field (see Table 1). Among them, professional knowledge refers to the specific domain knowledge and experiential knowledge possessed by an individual for a competent job position. It is mainly acquired through formal disciplinary education in schools, supplemented by social education, training, and other channels. Job skills refer to the professional techniques and comprehensive abilities that an individual

needs in a work position, and are important indicators for measuring whether one is competent for the job. They reflect an individual' s mastery of the occupation, including methodological ability and social ability. Methodological ability refers to the specific skills needed to meet work requirements, such as the ability to use office software; social ability refers to the ability to collaborate and communicate with others, such as communication and coordination ability. Personality traits refer to an individual' s persistent and stable responses to the external environment and information, including personality characteristics and basic individual traits. According to the HEXACO model, personality structure can be divided into six dimensions: honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, and openness to experience[26]. On this basis, personality traits are selected. Occupational literacy reflects an individual's cognitive attitude toward work and behavioral style, such as lifelong learning and teamwork spirit, and is closely related to factors such as education level and internship/employment experience.

Table 1 Indicator system of the job competency model in the library and information field

Indicator dimension	Competency indicators
Professional knowledge	D1 Library, information and archives management; D2 Computer science; D3 Human resources; D4 Mathematical statistics; D5 Marketing; D6 Information management; D7 Finance; D8 Law and economics; D9 Database knowledge; D10 Project management
Job skills	D11 Communication and coordination ability; D12 Decision-making and analytical ability; D13 Office software application ability; D14 Project management and implementation ability; D15 Learning ability; D16 Writing ability; D17 Emotional regulation ability; D18 Data analysis ability; D19 Logical thinking ability; D20 Software development ability; D21 Language expression ability; D22 Demand research and analysis ability; D23 Network programming ability; D24 English ability; D25 Leadership ability; D26 Information collection and analysis ability

Indicator dimension	Competency indicators
Personality traits	D27 Honesty-humility; D28 Emotionality; D29 Extraversion; D30 Agreeableness; D31 Conscientiousness; D32 Openness to experience; D33 Basic individual traits
Occupational literacy	D34 Teamwork spirit; D35 Overall awareness; D36 Service awareness; D37 Sense of responsibility; D38 Dedication; D39 Innovation awareness; D40 Organizational discipline; D41 Professional ethics; D42 Risk awareness; D43 Confidentiality awareness; D44 Dedication to research; D45 Lifelong learning; D46 Political literacy

4 Construction of a Job Knowledge Graph in the Library and Information Field Based on the Competency Model

This study adopts a top-down approach to construct a job knowledge graph for the library and information field, focusing on the processing of unstructured data. Based on the job competency model constructed above, knowledge is obtained from unstructured data; entity recognition, relation extraction, and knowledge fusion are completed using text-mining methods; and knowledge triples are then generated and stored together with structured data in the Neo4j knowledge-graph database, thereby completing the construction of the job knowledge graph for the library and information field.

4.1 Ontology Design

Ontology design is the foundational stage in constructing a knowledge graph. Common ontology construction methods include the METHONTOLOGY method, the skeleton method, the SENSUS method, and the seven-step method. This study selects the seven-step method, a semi-automatic ontology construction method with strong generality. The seven-step method includes seven iterative steps: determining the scope, reusing existing ontologies, enumerating terms, defining classes, defining properties, defining constraints, and creating instances. This study uses the ontology-editing software Protégé to create classes, relations, and properties. The specific steps are as follows.

as follows: First, the definitions of classes were preliminarily determined in combination with the text data obtained above, and classes were added in the “Classes” tab in Protégé. Second, based on the analysis results of job-position information, inter-entity relationships and attributes were added by clicking the

"Object Properties" tab in Protégé, and they were visually presented. The entity classes constructed in this paper and their relationships are shown in Figure 1, and a systematic description of the entity classes is given in Table 2.

Figure 1. Ontological framework of job information in the library and information domain

Table 2. Description of entity classes

Entity class	Description
Occupational category	Occupational categories classified according to the *Occupational Classification Dictionary of the People' s Republic of China (2022 Edition)*
Position name	Job-title information
Salary	Remuneration specified by the recruiting organization
Educational background	The applicant' s level of education
Work experience	The applicant' s professional experience and manifestation of abilities
Address	Address of the organization
Industry	Industry positioning
Recruiting organization	Name of the recruiting organization
Organization size	Number of employees in the recruiting organization
Organization nature	Institutional attributes of the recruiter, such as public institutions, state-owned enterprises, private enterprises, etc.
Professional knowledge	Theoretical knowledge and normative knowledge
Job skills	Core skills and competency requirements
Personality traits	A comprehensive manifestation of personality, habits, and values
Professional literacy	Recognition of professional values and compliance with ethical norms
Proficiency level	Level of mastery of job competencies

4.2 Entity Recognition

Before entity recognition, the initial data need to be sequence-labeled. Sequence labeling refers to labeling each character in the processed text so that entities can subsequently be identified. This study adopts the BIOES labeling method,

in which B denotes the beginning of an entity, I denotes the inside of an entity, O denotes a non-entity, E denotes the end of an entity, and S denotes a single-character entity. According to the job competency model constructed above, the entity categories in the dataset are divided into four types: professional knowledge (ZYZS), job skills (ZWJN), personality traits (GXTZ), and professional literacy (ZYSY). After code processing, the output data are text in txt format, in which each character occupies a separate line, and each line contains the character itself and its corresponding label, with the character and the label separated by a space.

This study selected Baidu ERNIE as the pre-trained model, mainly based on the following two considerations. First, ERNIE innovatively incorporates large-scale knowledge graphs into the pre-training process and strengthens deep understanding of Chinese entities, concepts, and semantic relationships through a continual learning mechanism, achieving excellent performance in semantic parsing tasks. Second, as a pre-trained model tailored to the characteristics of the Chinese language, ERNIE has significant advantages in handling Chinese lexical ambiguity, syntactic structures, and domain-knowledge integration, making it more compatible with the Chinese data scenario and analytical needs of this study. To facilitate parameter debugging at any time, model training was conducted in the Jupyter environment. Model parameters were saved after each training epoch so that, if training was interrupted, the code could be called from the interruption point to continue training. The ratio of the training set, validation set, and test set was $7:1:2$. After multiple rounds of training, among the actual 2,372 entities, the number of entities ultimately identified by the model was 2,087, of which 1,992 were correctly identified; the accuracy was 95%, the recall was 84%, and the F1 value was 89%. According to the experimental results, it can be judged that the model basically completed the entity-recognition task, and the experimental results were affected by the quality and quantity of the training set.

4.3 Relationship Extraction and Knowledge Fusion

In the relationship-extraction stage, the RoBERTa model was adopted. This model is optimized on the basis of the BERT model, uses a larger-scale training dataset and more batches, and increases the training duration so that the data can be trained more fully. The relationship-extraction process based on the RoBERTa model is specifically divided into three steps: Data input. The input layer is a high-dimensional representation composed of three low-dimensional vectors, namely character vectors, sentence vectors, and position vectors; "$" and "#" are used to mark the beginning and end of an entity, respectively. Text encoding. After preprocessing, the data are represented by the input-layer vectors and then undergo bidirectional encoding training through the Transformer. Generally speaking, when the amount of data is large, a larger number of Transformer layers can be selected; conversely, when the amount of data is small, the number of Transformer layers selected should also be smaller. After encoding

training, feature vectors fused with external features are obtained. Result output. The feature vectors obtained in the training stage are input into the fully connected layer, and a Softmax classifier is used to perform entity-relationship classification prediction. In the classifier, each input sample is calculated to obtain multiple sets of conditional probability values, the sum of each set of conditional probability values is 1, and relationship classification is performed according to the magnitude of the probabilities, taking the probability value

The entity relationship corresponding to the largest element is output as the model' s prediction result. During training, `batch_size` was set to 16, the maximum truncation length of the text sequence was 512, and the number of training epochs was 5. As with entity recognition, the model' s training performance was evaluated using a confusion matrix. After 5 epochs of training, the accuracy was 90%, the recall was 88%, and the F1 value was 89%. It should be noted that, in the actual extraction process, because some data express competencies rather vaguely, the effectiveness of relation extraction may be affected.

After knowledge extraction was performed on the unstructured text data, it was found that many degree terms were semantically redundant. If all of them were displayed in the knowledge graph, redundancy would result; therefore, degree terms with similar semantics needed to be fused. Specifically, cosine similarity was used to cluster degree terms, and degree terms falling within the same range were merged. The principle for calculating cosine similarity is as follows: in vector space, the closer the cosine value of the angle between two word vectors is to 1, the closer the angle is to 0°, indicating a higher similarity between the two word vectors. The text data processed into word vectors in the preceding step were imported into the `cosine_similarity` function for calculation. The degree term "proficient" was assigned a value of 1, and the similarity between the remaining degree terms and "proficient" was calculated. The results showed that the data exhibited obvious clustering intervals at 0.4, 0.3, 0.2, and 0.1, while the samples in the interval above 0.4 had a relatively large span, reaching as high as 1.0. Accordingly, the similarity values were divided into five intervals: $(0, 0.1]$, $(0.1, 0.2]$, $(0.2, 0.3]$, $(0.3, 0.4]$, and $(0.4, 1]$. The merged degree term for each range was determined using the median similarity value of that interval. Ultimately, the representative degree terms corresponding to the medians of the five intervals were "understand," "comprehend," "master," "skilled," and "proficient," respectively.

4.4 Knowledge Storage and Graph Generation

After entity recognition, relation extraction, and knowledge fusion of degree terms were completed, knowledge storage was carried out. Structured text was directly imported into the Neo4j knowledge graph database by reading tabular data, whereas unstructured text was processed through a series of steps to generate triple data, which were then imported into the graph database to generate nodes and relations; query functionality was implemented through the Cypher

language. The knowledge graph generated in this study contains a total of 15 nodes and 17 types of relations. Centered on job-position names and radiating outward, it can intuitively display the detailed information and required competencies of positions under each job category. Because the stored data volume is large and difficult to display in full, selected job-position information is used as an example (see Figure 2).

Figure 2. Example of a knowledge graph query for job-position information

The job-position knowledge graph presents the four dimensions in the competency model—namely, professional knowledge, job skills, personality traits, and professional qualities—as well as the keywords corresponding to the 46 indicators extracted above. Taking intelligence-related positions as an example, the four competency dimensions were queried using the Cypher language, yielding the following results (see Figure 3): in terms of professional knowledge, they include intelligence studies (corresponding to indicator D1), computer science (corresponding to indicator D2), and so on; in terms of job skills, mastery of office software application ability (corresponding to indicator D13), communication and coordination ability (corresponding to indicator D11), and so on is required; in terms of personality traits, they include conscientiousness (corresponding to indicator D31), and so on; in terms of professional qualities, team spirit (corresponding to indicator D34), and so on should be possessed.

Figure 3. Example of a competency graph for "intelligence service" positions

5 Design and Testing of an Application Platform for Job Knowledge Graphs in the Library and Information Field

To fully realize the value of job knowledge graphs in the library and information field, this study designs and constructs a corresponding application platform. Supported by the Neo4j knowledge graph database constructed above as a digital resource, the platform provides functions such as job search, job recommendation, and self-directed learning, aiming to offer job seekers employment paths guided by competency requirements.

5.1 Platform Architecture Design

At present, the B/S architecture is widely used in platform design and development. This architecture does not require each user to install a client program; it is convenient to use and highly sharable. Based on the B/S architecture, this study constructs an application platform for job knowledge graphs in the library and information field composed of a user layer, business layer, and data layer (see Figure 4).

Figure 4 Architecture of the Application Platform for Job Knowledge Graphs in the Library and Information Field

Visible labels in Figure 4:

- **User layer**
  - HTML
  - CSS
  - jQuery
  - Interface design: Vue.js framework
  - POST request
  - GET request
  - User interaction: Ajax
- **Business layer**
  - Job search
    - Search input content
    - Read database
    - Query visualization
  - Job recommendation
    - Construct graph
    - Similarity calculation
    - Generate recommendation list
  - Self-directed learning
    - Read database
    - Access external links
  - User management
  - System settings
  - Log records
  - ...
- **Data layer**
  - Neo4j
  - MySQL
  - Knowledge update

The user layer adopts the Vue.js front-end design framework for platform interface design, mainly including the platform' s usage structure, user-operation interaction, and visual effects. The technologies involved include HTML, CSS, jQuery, and others.

The business layer includes user function modules and system business modules. The user function modules include job search, job recommendation, and self-directed learning. In the job search function, after the user enters a query command in the browser, the platform rapidly retrieves information from the Neo4j knowledge graph database according to the command and feeds the retrieval results back to the browser, presenting the job knowledge graph in the display area. In the job recommendation function, after the user uploads a personal résumé, the back end constructs a personal résumé knowledge graph based on the résumé information, and can also match the personal résumé knowledge graph with job information in the database through a similarity algorithm to generate a job recommendation list. In the self-directed learning function, the

platform provides users with channels for improving personal abilities; users can click links to visit external websites and learn relevant job skills. The system business modules include user management, system settings, log records, and other services, which are used to maintain the normal operation of the website.

The data layer contains the Neo4j knowledge graph database and the MySQL database. The Neo4j knowledge graph database is used to store the job knowledge graph and provides data sources for job search and job recommendation. The MySQL database is used to store user information, external website links, and other content, and it is necessary to ensure that database resources are updated in a timely manner.

5.2 Functional Module Design

5.2.1 Job Search Function The multi-level and multidimensional query mechanism of the Neo4j knowledge graph database can enable intelligent job search. Applying the job knowledge graph to job search helps improve the user experience. This knowledge graph extracts key elements from job information and filters out invalid information from search results for users, thereby enabling them to grasp the core elements of jobs more quickly and satisfying the retrieval needs of different entities.

First, it satisfies the job-search needs of individual users. On the platform, users can query by entering keywords, such as job titles, major names, and recruiting organizations, and view the search results in the browser. For jobs of interest, users can click the job title to view detailed information. The job-information knowledge graph will be displayed below the interface, and users can view specific content by clicking individual nodes.

Second, it provides data references for talent-cultivation units such as universities in formulating talent-cultivation plans. Universities and other talent-cultivation units can, according to the employment-market demand data provided by the application platform, formulate and adjust talent-cultivation programs in a targeted manner, improve teaching quality, ensure that graduates possess the professional knowledge and skills required by the employment market, and cultivate high-quality interdisciplinary talent.

5.2.2 Job Recommendation Function As deep learning technologies have gradually been widely applied in recommendation systems, the level of intelligence of recommendation systems has also been continuously improving. Compared with traditional information recommendation, knowledge-graph-based information recommendation has the following advantages: first, it supports the formal semantic representation of massive heterogeneous distributed data; second, it supports intelligent association and reasoning to realize semantic user-interest modeling, item modeling, hybrid recommendation algorithms, and graph-based visualization of recommendation results [25]. At present, when online recruitment platforms conduct job recommendation,

most rely on the keywords searched by users and popular jobs as the basis. Users need to filter job information themselves on the basis of platform recommendations, while their personalized needs are ignored. Compared with the recommendation mechanism of traditional platforms, this application platform can identify the relationships between content and users, locate user needs more accurately, and also expand AI plug-in functions for collaborative filtering, thereby improving the platform' s computational efficiency.

The construction process of the personal résumé knowledge graph is similar to that of the job knowledge graph. A job seeker's résumé information includes age, educational background, work experience, industry, expected salary, expected work location, professional knowledge and skills possessed, and so on. After the user uploads the personal résumé to the application platform, the server uses entity recognition technology to annotate the entities in the résumé information, and then conducts relationship extraction to obtain triples similar to "Zhang San, educational background, bachelor' s degree" ,

relationship triple [Zhang San, proficient, data analysis], thereby completing the construction of the personal résumé knowledge graph. After the job-post knowledge graph and the personal résumé knowledge graph have been constructed, job recommendation is carried out through a similarity algorithm. In the knowledge-graph construction stage, semantic vector representations have already been generated for entities and relations. The Euclidean-distance algorithm is used to calculate the similarity between the entities and relations in the two types of knowledge graphs: the smaller the calculated result, the greater the similarity. Finally, the recommendation results are arranged in descending order of similarity, and the jobs recommended by the platform are displayed in the recommendation list.

5.2.3 Self-Learning Function

In addition to the two main functions of job search and job recommendation, the platform also designs an auxiliary function for self-directed learning. This function does not involve knowledge-graph technology; instead, it relies on the MySQL database and search-engine technology to perform keyword retrieval, with the aim of providing users with learning channels for the knowledge and skills that need to be mastered in the field of library and information science. The design of this functional module can, on the one hand, provide reference value for talent-training institutions such as universities in areas such as curriculum design; on the other hand, it can improve the core competitiveness of job seekers and practitioners, thereby enabling them to complete their work with higher quality.

5.3 Scenario Application and Functional Testing

To ensure the stable operation of the application platform for the job-post knowledge graph in the field of library and information science, the research team

recruited 25 library and information science students with employment needs to conduct functional tests in different scenarios.

5.3.1 Functional Test of the Job-Search Function

For the job-search function module, considering the diversity of online recruitment platforms and the coverage of the test sample, 5 students were randomly selected as the test group to use the application platform for the job-post knowledge graph in the field of library and information science to conduct job searches. Another 20 students served as the control group and were evenly divided into 4 groups to perform the same operations on four popular online recruitment platforms: "51job," "BOSS Zhipin," "Zhaopin," and "58.com." After completing platform registration, the 25 students conducted job searches simultaneously, and comparisons were made in terms of retrieval efficiency and the relevance of retrieval results. The job-search time was set at 10 minutes, and the time each student spent browsing each job and the job information were recorded.

Members of the test group entered the platform webpage and, after completing new-user registration, could log in to the main page. In the job-search function module, subjects could search for the required job information according to job keywords. Taking the job "intelligence analysis" as an example, the subject clicked "keyword search," and after the server completed its operation, a list of "intelligence analyst" jobs was obtained. The subject could click any job in the list, and the page would immediately display the information for that job and its knowledge graph (see Figure 5). This graph was consistent with the presentation results in the Neo4j knowledge-graph database. After completing the browsing of this job, the subject could randomly view other jobs in the list.

Figure 5 Functional test results for the job-search scenario

The search results and browsing time were recorded. The test results showed that, within the same period of time, among the 20 subjects using online recruitment platforms, the maximum number of job-information entries browsed was 31 and the minimum was 22; whereas among the 5 subjects using the platform developed in this study, the maximum number of job-information entries browsed was 46 and the minimum was 29. After all job information browsed by the 25 subjects was summarized, 3 practitioners in the field of library and information science were invited to conduct a professional relevance analysis of the job information. The analysis results showed that, on the four online recruitment platforms, the matching degrees between the job information browsed by the subjects and the professional job requirements were 75%, 81%, 56%, and 78%, respectively; whereas on the platform developed in this study, the matching degree between jobs and professional requirements reached

reached 90%. The test results show that the knowledge graph application platform developed in this study can improve users' efficiency in job searching, and that the retrieved job information is highly consistent with users' professional position requirements.

5.3.2 Functional Test of Job Recommendation

To test the job recommendation function of the knowledge graph application platform, participants were asked to upload their résumés in local mode. The server parsed the résumé information, constructed a personal résumé knowledge graph, and matched it with the job-position knowledge graph. As shown in Figure 6, the platform generates a job recommendation list in descending order of matching degree. After clicking the position with the highest matching degree, it can be seen that factors such as experience requirements, location requirements, and job competencies are highly consistent with the personal résumé information, indicating that the job-matching effect is good.

Figure 6 Functional test results for the job recommendation scenario

5.3.3 Functional Test of Self-Directed Learning

Participants entered the self-directed learning module for testing. As shown in Figure 7, in this module, the left-side functional area contains three items: Knowledge Energy Station, Skills Learning, and Workplace Competence, corresponding respectively to the three dimensions of professional knowledge, job skills, and occupational competence among the job competency elements. Participants can conduct self-directed learning according to their own needs. Since some personality traits tend to stabilize as an individual' s growth experience accumulates and are difficult to change through short-term social learning, the platform does not include them within the scope of self-directed learning. In this test, taking the learning of "data analysis" as an example, after clicking "Skills Learning," the platform provides relevant learning links, and the response speed is relatively fast.

Figure 7 Functional test results for the self-directed learning scenario

6 Conclusion

At present, the employment situation for university graduates is becoming increasingly severe. Empowering the precise matching of talent supply and demand in vertical fields through knowledge graphs has become a key pathway for promoting high-quality employment development. Addressing the problem of information asymmetry in job information retrieval in the library and information field, this paper constructs a job-position knowledge graph based on a competency model, covering 4 dimensions and 46 indicators, and designs an application platform that includes job search, job recommendation, and self-directed learning functions. On the one hand, this study helps job seekers accurately match job requirements and improve job-search efficiency; on the other hand, it provides decision-making references for talent-training institutions to optimize curriculum design and promote the integration of industry and education. In the future, the platform' s intelligent recommendation functions and learning service experience will be further optimized, promoting the transformation of

the job-position knowledge graph in the library and information field from being "searchable" to being "intelligently applicable."

References

[1] Central People' s Government of the People' s Republic of China. *Outline of the Fifteenth Five-Year Plan for National Economic and Social Development of the People' s Republic of China* [EB/OL]. [2026-03-15]. https://www.gov.cn/yaowen/liebiao/202603/content_7062633.htm?enable_bottom_share_style=1&hybrid_

[2] The CPC Central Committee and the State Council issued the *Outline of the Plan for Building China into a Leading Country in Education (2024–2035)* [EB/OL]. [2025-10-19]. https://www.gov.cn/gongbao/2025/issue_11846/202502/content_7002799.html.

[3] Cheng Jiejing, Wang Puyu, Shao Dan, et al. Discipline construction and development of library, information, and archives management in China—Reflections based on the High-End Forum on Library and Information Education and Discipline Development (2021) [J]. *Information Studies: Theory & Application*, 2022(2): 1-9.

[4] Xiao Yanghua, Xu Bo, Lin Xin, et al. *Knowledge Graphs: Concepts and Technologies* [M]. Beijing: Publishing House of Electronics Industry, 2020: 20-26.

[5] Shen T, Zhang F, Cheng J. A comprehensive overview of knowledge graph completion [J]. *Knowledge-Based Systems*, 2022: 109597.

[6] Abu-Salih B, Al-Tawil M, Aljarah I, et al. Relational learning analysis of social politics using knowledge graph embedding [J]. *Data Mining and Knowledge Discovery*, 2021(4): 1497-1536.

[7] Yu T. Research on the fusion of design and business knowledge graph for an intelligent decision-making system based on DSR theory [J]. *Scientific Journal of Intelligent Systems Research*, 2025(6): 66-73.

[8] Xu Tongyang, Huang Yingsi. Construction of a knowledge graph for celebrity chronological-biographical resources: A case study of Xu Shuofang' s *Chronological Biography of Late-Ming Dramatists* [J]. *Digital Library Forum*, 2022(12): 29-36.

[9] Xiong Yunlong. Construction of a knowledge graph for tourist attractions in Guizhou Province based on multi-source data fusion [C] // *Proceedings of 2022 2nd International Conference on Higher Education Development and Information Technology Innovation*. Hong Kong: Hong Kong New Century Cultural Publishing House, 2022: 281-282.

[10] Yang Yan, Lai Shuning. Research on knowledge graphs based on collaborative knowledge-construction conversations [J]. *Modern Educational Technology*, 2025(8): 97-106.

[11] Zheng Chunrong, Ning Yunjing. Research on knowledge graphs supporting the disciplinary development of area studies [J]. *Journal of Northwest University (Philosophy and Social Sciences Edition)*, 2025(6): 94–105.

[12] He Chunhui, Guo Boqian. A job matching and ranking method based on knowledge graphs and semantic similarity [J]. *Journal of Hunan City University (Natural Science Edition)*, 2021(5): 59–63.

[13] Zhang Xinxin. Research and application of knowledge graph construction for data-related positions based on a competency model [D]. Changchun: Jilin University, 2022.

[14] Sun Yi, Yuan Hangping, Zheng Li, et al. A review of knowledge-enhancement methods for natural language pre-trained models [J]. *Journal of Chinese Information Processing*, 2021(7): 10–29.

[15] Briskila J, Subalalitha C N. An ensemble model for classifying idioms and literal texts using BERT and RoBERTa [J]. *Information Processing & Management*, 2022(1): 102756.

[16] Analysys. Insights into the online recruitment and job-seeking industry in 2023 [EB/OL]. [2025-10-01]. https://www.analysys.cn/article/detail/20021133.

[17] iResearch Consulting. 2024 China online recruitment industry research report [EB/OL]. [2025-11-27]. https://report.iresearch.cn/report_pdf.aspx?id=4701.

[18] Revision Working Committee for the *National Occupational Classification Dictionary. Occupational Classification Dictionary of the People' s Republic of China: 2022 Edition* [M]. Beijing: China Human Resources and Social Security Publishing Group, 2022.

[19] An Ran, Chu Jihua, Hong Xianfeng. Research on a framework for intelligence analysis methods oriented toward unstructured data [J]. *Information Studies: Theory & Application*, 2024(2): 143–150.

[20] Liu Baoxia, Gu Yuqi. Based on the competency model: Making organizational performance both attractive and practical [J]. *Human Resources*, 2023(23): 20–22.

[21] Li Hai. A review of research on competency models [J]. *Journal of State Grid Technology College*, 2020(4): 27–32, 45.

[22] Wang Dezhuang, Lin Mengmeng. Research on constructing a competency model for data-related positions for Master of Library and Information Science students: Data analysis based on enterprise recruitment websites [J]. *Information Exploration*, 2024(6): 1–9.

[23] Liu He, Wang Lili. Construction and improvement paths of a competency model for librarians engaged in reading promotion [J]. *Library Science Research*, 2025(12): 59–70, 91.

[24] Wang Jiahui. Research on constructing a competency model for cultural and creative librarians in public libraries in China [D]. Nanning: Guangxi Minzu University, 2025.

[25] Zheng Ze. Research based on the Word2Vec word-embedding model [D]. Fuxin: Liaoning Technical University, 2018.

[26] Zhang Wei, Wang Fei. Prospects for research on the HEXACO personality traits model [J]. *Journal of Lanzhou University of Arts and Science (Social Sciences Edition)*, 2016(5): 65–72.

Construction and Application of a Competency Model-Based Knowledge Graph for Positions in the Library and Information Science Field

Shi Yali1 Liang Miao2 Xiong Xinni1 Zhang Lixia1

1. School of History and Culture, Hubei University, Wuhan, 430062

2. Huanggang Polytechnic University, Huanggang, 438000

Abstract To address the challenges of information isolation and asymmetry in retrieving job information in the library and information science (LIS) field, this paper attempts to construct a knowledge graph for LIS positions based on a competency model and to design and develop the architecture and functionality of its application platform. The study initially uses online recruitment information in the LIS field as raw data. Using tools such as Python, it conducts cluster analysis on job postings and categorizes job competencies in this field into four dimensions: professional knowledge, job-specific skills, personal traits, and professional ethics, encompassing 46 competency indicators, including decision-making analysis, data analysis, and language expression. Second, Protégé is employed to create an ontology, and a text-mining approach is adopted to construct the knowledge graph for LIS positions. Finally, an application platform for the knowledge graph of LIS positions is designed and tested for its scenario-based functionalities. Test results show that this application platform can effectively improve job seekers' information retrieval efficiency and holds decision-making reference value in optimizing job recommendation services on recruitment platforms and promoting innovation in talent cultivation models.

Keywords Job of library and information science; Competency model; Knowledge graph; Text mining; Application platform

Note: Figure translations are in progress. See original paper for figures.