SPECIAL SECTION ON GEOMETRIC ALGEBRA IN SIGNAL PROCESSING

The General Higher-Order Neural Network Model and Its Application to the Archive Retrieval in Modern Guangdong Customs Archives

MENG WANG\textsuperscript{1} AND LILAN CHEN\textsuperscript{2}
\textsuperscript{1}School of Information Management, Sun Yat-Sen University, Guangzhou 510275, China
\textsuperscript{2}School of Foreign Languages, Guangdong Pharmaceutical University, Guangzhou 510006, China
Corresponding author: Lilan Chen (chenllan@mail2.sysu.edu.cn)

ABSTRACT
Because of the unique attributes of archive information, it is challenging to manage and effectively retrieve archive information in the archive information management practice. This paper designs and develops the first general higher-order Neural Network Model for archives. Based on the analysis of the correlation, the relevance of the weight model, the study of technical methods about the core weight, the direction weight retrieval, and the statistical ranking of the results, this paper designs a corresponding archive information analysis system. Finally, this paper adopts the B/S development model by applying the relevance ranking weight algorithm into the comprehensive archive retrieval activities, which not only enhances the intelligence and efficiency of the archive retrieval, but also can act as a standard example to demonstrate informatization construction for archive management. This paper compares this algorithm with two other existing retrieval algorithms and verifies the practicability of the relevance algorithm by evaluating the algorithm and the default retrieval algorithm using the NDCG evaluation method.

INDEX TERMS
Higher-order neural network model, core weight, direction weight, evaluation method, archive information, modern Guangdong customs archives.

I. INTRODUCTION
Information retrieval is a process and a technique that allows information users to find relevant information they need through a dataset, where information is organized in a certain way [1]–[3]. A document retrieval system contains mainly three parts: index generation, query processing, and document retrieval. The information retrieval field, where correlation [4]–[6] is a core research topic, aims at information filtering. The relevance of the information retrieval means a matching relationship between the textual content and the queries in the information source, and the matching relationship is multi-dimensional, dynamic, complex, and measurable. Fuzzy retrieval, also known as “containing search”, is a retrieval approach that takes searching character strings or parts of textual content in the document as searching keywords [7]. This approach can expand the retrieval range, thus achieving larger search result sets. The fuzzy retrieval approach, which uses keywords as searching queries, has improved the usability and user-friendliness of the system from the user perspective, especially when there is a large amount of retrieval content.

The archive description and its indexing [8], [9], the basic information in archive information retrieval, are the foundation of retrieval technology. The process of archive description [10]–[12] is to extract the required index information and record it. The archive description should be in line with the principles that were written in the Archive Description Rules, which is formulated by the Chinese government and is used to standardize the archive description work. The archive indexing is to analyze and select the context and themes of the archives during the archive description process and provide a standardized language through concept conversion. The indexing reveals the classification attributes and theme of the archives, providing an approach to archive retrieval from these two aspects.

Regarding retrieval methods, a large amount of related foreign literature has focused on improving the retrieval performance by providing methods to organize the information sources. These methods can be classified into two dimensions: the construction of retrieval tools and the expansion of semantic information. For the construction of retrieval tools,
tools, Silvia S. K. [13] has discussed the problems faced by the Israel Archives, including constructing the retrieval tools and comprehensive dictionaries. Silvia discussed series of methods to search information from metadata and content of archives by using the thesaurus index. Based on the study of Silvia, this paper proposes constructing the archive description and information retrieval system on the basis of ISAD and ISAAR. Niu J. F. [14] sought to facilitate the retrieval of information objectives in certain institutions, such as culture heritage institutions. To do so, Niu analyzed the differences between events and function based on the event-based information organizational method; discussed the uses of events as a source to organize and describe archive information; and redesigned two methods for describing the metadata in archives. The expansion of semantic technique has always been one of the hot topics in information retrieval study.

Milne C. [15] studied the adaptability of context classification of archive retrieval in the portal or internal network development, which aims to establish a stronger interdisciplinary relationship in the entire information industry and promote the development of information retrieval discipline. Bak G. [16] has pointed out the drawbacks of record classification of the electric record system and appealed to expand the record classification definition by capturing semantic information expansion for archive sources, breaking constraints of paper-based archive retention rules, improving the efficiency of information retrieval, realizing the project-level management of recording, preserving archive documents, and transforming archive practices. Ricardo E. B. [17] has proposed a collaborative framework for archive information system based on the extensible markup language EAC-CPF (Coded Archive Context). This framework uses EAC-CPF to promote users’ interactive experiences with internet on the basis of EAC-CPF’s features to share the context and authority records; support auxiliary navigation and topic mapping; and provide a semantically rich access layer to ensure the location of different archives. Machin J. [18] has reviewed Stewart’s book Practical Ontology for Information Professionals about the ontology theory, including the adoption, construction, and retrieval of this method and its application in semantic contexts. The ontology method presented in this book is a good guidance for archive work in both government and institutions.

The aim of information retrieval is to offer the search results to information users. However, due to the huge retrieval content, the information user cannot go through all the results, contrary to the user-friendly principle for software development. This paper constructs a Neural Network Model for archive information retrieval, which not only can support users’ self-retrieval by simplifying the searching queries, but also can efficiently filter redundant information that affects the readability of the retrieval results. Currently, due to a lack of well-designed mechanics of statistical results ranking in archive information systems, information users have no methods to deal with significant numbers of retrieval results, thus affecting their work efficiency. Therefore, this paper proposes a higher-order Neural Network Model to support Neural Network self-learning [19], [20], which creates the self-adjustment of retrieval condition and requirement consistent with self-learning needs. This paper includes four parts: the first part explains the motivation to construct the New Incentive Neural Network Model; the second part illustrates the creation of theoretical propositions of the Neural Network Model construction; the third part describes and analyzes the archive case study by applying this proposed model; and the fourth part compares this algorithm with two other existing retrieval algorithms. This verifies the practicability of the relevance algorithm by evaluating the algorithm applied in this paper and the default retrieval algorithm through the Normalized Discounted Cumulative Gains (NDCG) evaluation method.

II. THE CONSTRUCTION OF GENERAL HIGHER-ORDER NEURAL NETWORK MODEL FOR ARCHIVES

A. HIGHER-ORDER NEURAL NETWORK MODEL STRUCTURE (HONN)

The structure diagram of the higher-order Neural Network Model is as follows:

The structure of the general higher-order neuron network model [7], [8] can be a three-layer forward network with a single hidden layer.

The output of the \( i \)-th neuron in the \( K \)-order neural network is:

\[
Y (i) = f_i [T_0 (i) + T_1 (i) + T_2 (i) \ldots + T_k (i)] \tag{2.1}
\]

Here, \( f_i \) is the activity function of the \( i \)-th neuron, \( K \) is the highest power of the weight sum of neuron inputs, called \( K \)-order, and \( T_k (i) (k = 1, 2 \ldots, k) \) is the \( k \)-th term of the \( i \)-th neuron. It is the product weight of the \( k \) components of the input vector. That is:

- 1\(^{st}\) -order: \( T_1 (i) = \sum_j W_1 (i, j) X (j) \)
- 2\(^{nd}\) -order: \( T_2 (i) = \sum_j \sum_l W_2 (i, j, l) X (j) X (l) \)
- 3\(^{rd}\) -order: \( T_3 (i) = \sum_j \sum_l \sum_m W_3 (i, j, m, l) X (j) X (l) X (m) \)

where \( X (j) \), \( X (l) \), and \( X (m) \) are the \( j \)-th, \( l \)-th, and \( m \)-th components of the input vector for the \( i \)-th neuron. \( W_3 (i, j, m, l) \)
is the corresponding joint weight which represents the correlation between the product of k inputs and the output, while \( T_0(i) \) is the threshold value of neurons.

Similar to the multilayer perceptron, this higher-order neural network has strong nonlinear transformation characteristics. Using Volterra series theory, it can be proved that a higher-order neural network suitably high enough can realize arbitrary nonlinear transformations within the precision range.

**B. THE STRUCTURE OF THE GENERAL HIGHER-ORDER NEURAL NETWORK MODEL**

The structure of the general higher-order Neural Network Model is similar to that of the RBF network model. The structure is shown in Fig. 2, where the input layer is composed of information source nodes, the second layer is the hidden layer, and the number of units depends on the problems described. The third layer is the output layer, responding to the function of the input mode. It can be replaced by a decision function or it can adopt linear or nonlinear functions. The transformation from the input layer space to the hidden layer space is nonlinear, as shown in the following formula:

\[
O_i = \sum_{j=1}^{n} \left( \frac{W_{ji}}{|W_{ji}|} \right)^{2} \left( |W_{ji}(X(j) - W_{ji}^{'}| - \theta \right) (2.2)
\]

This is a kind of non-negative, nonlinear function with symmetric distribution around the center point.

The structure of a single higher-order neuron is shown in Fig. 2. The neuron inputs the R dimension, letting \( P = (P_1, P_2, P_3, ... , P_R)^T \) represent the center weight vector, \( W' = (W'_1, W'_2, W'_3, ... , W'_R)^T \). The direction weight vector is \( W = (W_1, W_2, W_3, ... , W_R) \), where the subscript 1 represents the first neuron. “Dist” is a concept of generalized distance. Obviously \(|dist| = \sum |(P - W') \ast W'|^{N} \), where \( \theta \) represents threshold, equivalent to the distance from the point on the hypersurface to its center, the excitation function is the hard limiting function.

![FIGURE 2. The structure model of general higher-order neural network model for archives.](image)

The general higher-order Neuron network model has the following features:

1. The input nodes can be multi-dimensional and the network structure is a fixed three-layer network (including the input layer).
2. The neurons in the higher-order neural network adopt a general calculation formula. This formula consists of parameters representing different meanings. Each neuron can choose different parameters according to different needs to display different hypersurface shapes in the multi-dimension space.
3. Neural networks are no longer composed of a single-neuron model. The same neurons can form specific functional modules to solve specific problems, and several functional modules can form a complex neural network to solve complex problems. This is like the different shapes and functions of biological nerve cells that make the construction of neural networks more flexible and convenient.
4. Higher-order neurons are usually locally sensitive only to the input space. This means only when the input vector falls in a specific region of the input space can the higher-order neurons produce a non-zero response.
5. Higher-order neurons perform a nonlinear transformation on the generalized distance between the input and its center. This nonlinear transformation is adjustable.

Archive retrieval includes retrieval of different levels, such as volume level retrieval and document level retrieval. This research selects series of different types of archive documents as examples to illustrate the process of higher-order information transformation by using the document level retrieval method.

**Higher-order Neural Network Model and Learning for Archive Retrieval**

**Step 1:** Input R number of PDF documents as samples, named as archive document 1, archive document 2, ..., archive document R. These documents go through the understanding and retrieval stages in the hidden layer, but these analysis and retrieval results are ready to be affected by kinds of things such as the employees’ foreign language capabilities, especially their English skills; their historical knowledge, especially their relevant social, political and economic knowledge in the period of late Qing to the Republic of China; their professional skills, especially about archive description; the standard of archive description; knowledge of archive classification; whether the archive information is consistent with the archive content; and language expression skills for archives.

**Step 2:** The employees will examine and verify the results; only the verified documents can enter the output layer. These relevant documents will be renamed as document 1’, archive document 2’, ..., archive document R’. The corresponding center weight is constructed as \( W' = (W'_1, W'_2, W'_3, ... , W'_R)^T \).

**Step 3:** To accurately retrieve archive documents, it is necessary to correctly understand the archive documents. Then, one can proceed with the searching process according to the retrieval principles and standards, in which the employees’ competence level acts as a parameter to affect the information retrieval process and results.
Step 4: During the information retrieval test, the participants are divided into different groups to do different tasks. Set \( W = (W_1, W_2, W_3, ..., W_R) \) to represent the participants’ competence levels for understanding archive document 1, archive document 2, ..., archive document \( R \). The participants’ duties are to understand and markup these archive documents. These participants have different competence levels; therefore, the results of information retrieval are affected by their levels of understanding. This is concerned with the first stage of the hidden layer—understanding skills.

Step 5: Set \( (W_1, W_2, W_3, ..., W_R) \) to represent different factors that affect the ability to understand, the auditing standards determine the direction weight of its network \( W = (W_1, W_2, W_3, ..., W_R) \).

Step 6: Following the process of understanding the content of archive documents is the archive retrieval process. Set \( L_1, L_2, ..., L_m \) to represent employees’ cognition levels of archive documents. These employees have different cognition levels about the requirements, elements, and principles of archive retrieval. We can verify this from the second stage of the hidden layer.

Step 7: These archive documents are examined by standard retrieval rules, and we set \( LH_1, LH_2, ..., LH_p \) to represent specific principles and requirements of archive retrieval, \( O_1, O_2, ..., O_p \) to represent employees’ capabilities of inspecting and auditing the retrieval quality of archives, which are also influenced by many factors.

Step 8: Use \(|dist| = \sum |(P - W') \times W| \) to motivate \( y = C(O_1, O_2, O_3, ..., O_m) = y_s \). If these documents are being qualified, then the archive document \( 1' \), archive document \( 2' \), ..., archive document \( R' \) will be formalized.

III. COMPARATIVE ANALYSIS OF THE REALIZATION PROCESS OF ARCHIVE INFORMATION RETRIEVAL

A. THE DATABASE

The Guangdong Provincial Archives has a large collection of Guangdong Customs archives in modern times, which were formed between 1861 and 1949, with a total of 16,115 volumes and a total of 3.87 million pictures [21]. In this experiment, we first extract 10 volumes of modern Guangdong Customs archive documents containing 2,572 archive documents: 129 in the secretary category, 261 in the personnel category, 597 in the customs category, 453 in the investigation category, 1,490 in the trade category, and 287 in other categories. Their composition is shown in Fig. 3. The experiment made classification according to each type of keywords: the secretary category has the keyword “secretary”; the personnel category has the keyword “appointment”; the tax category has the keyword “preferential tax”; the investigation category has the keywords “smuggling and prevention policemen”; the trade category has the keywords “opium, munitions, and tea.” Through the analysis of the NDCG comprehensive index of the keywords, they were added and classified according to the comprehensive index.

B. CALCULATION OF THE WEIGHT VALUE

Some of these archive documents are selected as examples to do information retrieval experiments. The weight value depends on the input searching orders. For instance, if the user types in the keywords “for future reference”, “registration” in a sequential order, then, “for future reference” has the maximum weight value, next turns to the second input keyword, and the “registration” has the minimum weight value due to its last input order. We call this weight the core weight, \( W = (W_1, W_2) \). According to the statistic of the query log about the total query numbers and query numbers of the two keywords, the query numbers’ weight values of “for future reference” and “registration’” are 0.05 and 0.03 respectively. This is called as direction weight \( W' = (W'_1, W'_2) \), because all the archive record information is stored in “archive record table”, which has 10 fields. Therefore, query keywords can appear in the 10 fields and are ranked by the importance degree of the fields. The weight value of each field can be calculated based on the location weight value, where each keyword appears in these 10 fields. The direction weight is formed based on the historical statistic numbers of the two keywords in each field, which are calculated by the ratio between the importance weight of keywords and the weight of the historical query times.

C. COMPARATIVE ANALYSIS

The NDCG evaluation method assigns different scores (0-K) to the relevance of each search result, \( k \) can be set in advance based on actual needs. The score of the relevance is determined subjectively. Generally, it is evaluated collectively by some experienced appraisers, and its definition is shown in formula (3-1).

\[
N_q = M_q \sum_{j=1}^{m} (2^{r(j)} - 1) / \log(1 + j)
\]

In this formula, \( N_q \) is defined as the ordering value, \( M_q \) is the standardized parameter, \( r(j) \) is the relevance value of the result at location \( j \), \( q \) is the query efficient, \( m \) is the number of the returned results, and \( j \) is the record number.

The NDCG method takes position as an influential factor of the relevance, \( r(j) \) represents the overall value between values (0 to k) achieved by using the higher-order neural network at position \( j \). As formula (3-1) displays, \( r(j) \) represents...
TABLE 1. All indicators of experimental archive documents.

| classification | Keywords       | CG[21] | DCG[22] | NDCG[23] | Ours |
|----------------|----------------|--------|---------|----------|------|
|                |                | KR     | CR      | KR       | CR   | KR   | CR   |
| Secretary      | Appointment    | 57.1   | 57.1    | 64.1     | 64.1 | 65.2 | 65.2 |
|                | tax            | 79.2   | 66.7    | 82.4     | 70.1 | 85.6 | 75   |
|                | Preferential tax| 59.5   |         | 57.5     | 71.2 | 82.4 |
|                | Tax-free       | 61.4   |         | 70.3     | 67.2 | 69.2 |
| Investigation  | Smuggling      | 71.2   | 80.9    | 75.3     | 84.9 | 88.1 | 89.3 |
|                | Policemen      | 90.5   |         | 94.5     | 92.4 | 93.7 |
| trade          | Opium          | 59.5   | 64.9    | 60.4     | 67.5 | 64.5 | 69.7 |
|                | Arms           | 56.7   |         | 62.1     | 63.4 | 71.9 |
|                | Tea            | 78.4   |         | 80.1     | 81.4 | 90.8 |
| Personnel      | Appointment    | 87.1   | 87.1    | 81.2     | 81.2 | 67.3 | 67.3 |

Note: KR=keywords rates, CR=classification rates

the archive document scores which are provided by people: 0 represents the best retrieval, k represents the worst retrieval. In this way, r(j) is a discrete numerical sequence. Performing exponential operation on r(j), there is a significant difference of the results within different calculations of r(j). Therefore, the r(j) values at each location have a dramatic impact on the evaluation results. When using NDCG to do evaluation, it is required to use the formula to calculate each query q within the retrieval set, and get the average value of all the NDCG retrieval results.

D. NDCG COMPREHENSIVE INDEX ANALYSIS

The two ideas of Discounted Cumulative Gains (DCG) are as follows. First, high relevance results are much more influential on the final index score compared with the general relevance results. Second, if the high relevance results appear at the top position, the index will be higher. This paper adds comparative analysis of corresponding indicators.

Cumulative Gains (CG) is the predecessor of DCG, which includes the correlation degree of the relevance but excludes the position factors. It is the sum of the correlation scores of the search results. The CG at the designated position q is:

$$CG_q = \sum_{i=1}^{q} rel_i$$  (3.2)

where rel_i represents the relevance degree at position i.

DCG divides the result of each CG by a loss value. It aims at making the higher-ranking results more influential on the final results. Assume that the lower the order is sorted, the lower the value is. At position i, the value is 1/log_2(i+1), thus the benefit from the i-th result is rel_i * 1/log_2(i+1) and the DCG indicator is:

$$DCG_q = \sum_{j=1}^{m} rel_i / \log_2(i+1)$$  (3.3)

When using NDCG the retrieval results change based on different keywords; thus, the number of returned results is not consistent. The DCG is an accumulated value, and cannot compare two different retrieval results; therefore, it needs to be normalized via division by IDCG:

$$nDCG_q = \sum_{j=1}^{m} DCG_q / IDCG_q$$  (3.4)

Here, IDCG_q = \sum_{j=1}^{j(REL)} (2^{rel_i} - 1)/\log_2(i+1), |REL| represents, by the set of the top q results, the relevance order from the largest to the smallest order. That is to say, the results are sorted in an optimal manner.

We will do preliminary archive description and classification experiments by analyzing, selecting, and recording the content and features of different types of archives in the modern Guangdong Customs archives, such as classifications of secretary, personnel, tax, investigation, trade, and other classifications. The classification of the archives is shown in Table 1.

According to the calculation index of each archive document, we get the archive classification as shown in Table 2.

Table 2 shows that for the classifications of secretary, trade, and personnel archives, we used the proposed method to calculate the direction weight value of secretaries, taxes, trade, and personnel archive categories. The results are presented as

$$(W_{11}, W_{21}, W_{22}, W_{23}, W_{31}, W_{32}, W_{41}, W_{42}, W_{43}, W_{51})$$.

W_{11} is the direction weight value of appointment for the keywords of the secretaries category; W_{21}, W_{22}, W_{23} are the direction weight values for the keywords of the tax, preferential tax, and tax exemption for the tax category; W_{31}, W_{32} are the direction weight values for the keywords of the smuggling
TABLE 2. Experimental archive classification.

| classification | CG[21] | DCG[22] | NDCG[23] | Ours |
|----------------|--------|--------|--------|------|
| Secretary      | 73.5   | 75.1   | 90.5   | 91.2 |
| Tax            | 62.4   | 67.9   | 71.8   | 81.4 |
| Investigation  | 59.3   | 70.1   | 79.7   | 79.2 |
| Trade          | 70.9   | 88.9   | 85.2   | 93.5 |
| Personnel      | 81.7   | 79.8   | 79.9   | 82.7 |

and prevention police for the investigation category; \( W_{41}, W_{42}, W_{43} \) are the direction weight values for the keywords of the opium, munitions, and tea for the trade category; and \( W_{51} \) is the direction weight value of appointment for the keywords of the personnel category. Ours(1) presented the first results of direction weight value randomly; thus, the classification was not satisfactory. With 10 times of adjustments, we achieved Ours(10) and found the results representing the secretaries, tax, trade, and personnel categories calculated by using the proposed method are better than the results calculated by DCG and other extended methods. Although the results representing investigation category by using the proposed method were slightly lower than those using NDCG [23], they were higher than the results presented by other gain indicators. After 50 times of adjustments, Ours(50) indicates that the proposed method is better than DCG and other extended methods in secretaries, tax, trade, and personnel categories.

E. EXISTING PROBLEMS

Although this paper has verified that the proposed method is better than DCG and other extended methods in solving the archive classification issues in archive description, there are still some drawbacks:

1. This proposed method is based on keywords to do classification; therefore, if the classification keywords are not constructed well, the results of archive classification are not as good as expected.

2. This proposed method is based on the learning algorithm of the direction weight value. The constructed motivation function \(|\text{dist}| = \sum |(P - W') \ast W|^N\) generates the optimal classification model \( y = C(O_1, O_2, O_3, \ldots, O_m) = y_s \). Therefore, the index \( N \) influences the results of the classification. However, for this study, we only made the experiment setting of \( N \) be 2. We may need a further discussion on other settings of the index \( N \) to make sure the classification is optimal.

3. This proposed method is concerned with the core weight learning algorithm, while the core weight is the key point to construct the center point of the classification model. This paper adopts the formula \( W(t) = W(t-1) - \eta \lambda \) to adjust the core weight of classification. The parameters \( \eta, \lambda \) in the formula play a recursive decline, so that the learning algorithm can reach a relative minimum, but how to achieve the minimum in order to realize optimal archive classification still needs to be addressed.

These problems mentioned above are the future work for the method proposed in this paper.

IV. CONCLUSION

This paper first did an analysis on existing retrieval technologies, analyzing their advantages and drawbacks. Secondly, by doing an in-depth review on retrieval technology, this study found the relevance of keywords plays an important role in ranking information retrieval results. Therefore, this paper takes advantage of the basic relevance principles of the higher-order neural network to expand a new model for information retrieval technology. This model proposed a modification algorithm based on calculating the core weight and direction weight of the higher-order neural network for archive users, efficiently solving the query structure problems of information retrieval, optimizing the retrieval process, and improving the efficiency of information acquisition. Then, this paper applied relevance algorithms to the comprehensive searching activities of the modern Guangdong Customs archives, enhancing the intelligence and efficiency of the archive retrieval process, thereby providing a good example for developing information construction of archive information management. Finally, the correlation ranking algorithm in this paper was compared with two other existing sorting algorithms. The NDCG evaluation method was used to evaluate the proposed algorithm and the comparison algorithm, which verified the efficiency of the correlation algorithm proposed in this paper. The experimental process shows that the algorithm is correct and efficient. The research on archive information retrieval techniques has highly practical value. This paper has provided detailed analysis and in-depth discussion of information retrieval technologies and applied the research output (i.e., the new model) into a case study to verify its performance. However, whether from the perspective of information retrieval study or from the perspective of the importance of archive information management research, there is still much room for expanding the retrieval technology for archive information systems.

REFERENCES

[1] T. Gagie, A. Hartikainen, K. Karhu, J. Kärkkäinen, G. Navarro, S. J. Puglisi, and J. Sirén, “Document retrieval on repetitive string collections,” Inf. Retr. J., vol. 20, no. 3, pp. 253–291, Jun. 2017.
[2] W. Cao, Q. Liu, and Z. He, “Review of pavement defect detection methods,” IEEE Access, vol. 8, pp. 14531–14544, 2020.
[3] D. Cobas and G. Navarro, “Fast, small, and simple document listing on repetitive text collections,” in Proc. Int. Symp. String Process. Inf. Retr. Cham, Switzerland: Springer, Oct. 2019, pp. 482–498.
[4] N. Yameng, “Research and implementation of cross-language information retrieval technology,” M.S. thesis, College Comput. Sci. Technol., Xidian Univ., Xi’an, China, 2011, pp. 3–5.
[5] W. Jiayi, “Research on keyword fuzzy query and result set sorting strategy based on relational database,” M.S. thesis, College Comput. Sci. Eng., Central South Univ., Changsha, China, 2008, pp. 3–4.
[6] W. Cao, Q. Lin, Z. He, and Z. He, “Hybrid representation learning for cross-modal retrieval,” Neurocomputing, vol. 345, pp. 45–57, Jun. 2019.
[7] L. Chunxiao, M. Jun, and M. Xiangfu, “Classification method of fuzzy results based on membership degree,” J. Liaoning Univ. Technol., vol. 31, no. 5, pp. 2–3, 2011.

[8] D. H. Kraft and E. Colvin, “Fuzzy information retrieval,” Synth. Lect. Inf. Concepts, Retr., Services, vol. 9, no. 1, p. 1-63, Jan. 2017.

[9] S. J. Putra, M. Z. Na’fan, and M. N. Gunawan, “Integrated passage retrieval with fuzzy logic for indonesian question answering system,” Int. J. Perceptive Cognit. Comput., vol. 5, no. 2, pp. 31–34, Dec. 2019.

[10] M. Mühlung, M. Meister, N. Korthage, J. Wehling, A. Höirth, R. Ewerth, and B. Freisleben, “Content-based video retrieval in historical collections of the German broadcasting archive,” Int. J. Digit. Libraries, vol. 20, no. 2, pp. 167–183, 2019.

[11] M. Harris, M. Levene, D. Zhang, and D. Levene, “Comparing ‘parallel passages’ in digital archives,” J. Document., vol. 76, no. 1, pp. 271–289, 2019.

[12] K. J. Rawson, “The rhetorical power of archival description: Classifying images of gender transgression,” Rhetoric Soc. Quart., vol. 48, no. 4, pp. 327–351, Aug. 2018.

[13] S. Schenkolewski-Kroll and A. Tractinsky, “Archival description, information retrieval, and the construction of thesauri in Israeli archives,” Archival Sci., vol. 5, nos. 2–4, pp. 69–107, Dec. 2005.

[14] J. Niu, “Event-based archival information organization,” Archival Sci., vol. 15, no. 3, pp. 315–328, Sep. 2015.

[15] C. Milne, “Taxonomy development: Assessing the merits of contextual classification,” Records Manage. J., vol. 17, no. 1, pp. 7–16, Feb. 2007.

[16] G. Bak, “Continuous classification: Capturing dynamic relationships among information resources,” Archival Sci., vol. 12, no. 3, pp. 287–318, Sep. 2012.

[17] R. Eito-Brun, “Context-based aggregation of archival data: The role of authority records in the semantic landscape,” Archival Sci., vol. 15, no. 3, pp. 217–238, Sep. 2015.

[18] J. Machin, “Practical ontologies for information professionals,” Arch. Manuscripts, vol. 45, no. 1, pp. 55–56, Jan. 2017.

[19] W. Shoujue and L. Jiangliang, “Geometrical learning, descriptive geometry, and biomimetic pattern recognition,” Neurocomputing, vol. 67, pp. 9–28, Aug. 2005.

[20] W. Shoujue, S. Jingpu, C. Chuan, and L. Yujian, “Direction-basis-function neural networks,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), vol. 2, 1999, pp. 1251–1254.

[21] T. Leelanapab, G. Zuccon, and J. M. Jose, “A query-basis approach to parametrizing novelty-biased cumulative gain,” in Proc. Conf. Theory Inf. Retr. Berlin, Germany: Springer, 2011, pp. 327–331.

[22] G. Dupret, “Discounted cumulative gain and user decision models,” in Proc. Int. Symp. String Process. Inf. Retr. Berlin, Germany: Springer, 2011, pp. 2–13.

[23] A. Mishra and K. Berberich, “Leveraging semantic annotations to link wikipedia and news archives,” in Proc. Eur. Conf. Inf. Retr. Cham, Switzerland: Springer, 2016, pp. 30–42.