

Study on University Archives Information Based on Grid Computing and Context Perception

Xu Yiwu¹, Fu Wenjing²

¹Shandong Agriculture and Engineering University, Jinan Shandong, 250100, China

²Wuhan University of Science and Technology, Wuhan Hubei, 430000, China

Keywords: grid computing, context awareness, archive information

Abstract: With the continuous expansion of information resources, higher requirements are put forward for the archives management of higher education institutions. According to the requirements of scientific management of college archives management, a context-aware technology-based university archive index semantic engine in the paper is proposed for addressing the growing number of archives management categories and projects, and the classification of archives, the system data is processed by using grid computing. For fuzzy conceptualization of indexing requirements, the reference effect of the materialized index feature vector in the multi-dimensional initial set is studied, and the principle of optimization definition is presented. Finally, in the research and analysis of the research content, the experimental proves that the content of the paper has played an effective role in the information management of college archives, which provided a good technical support for the process of information of archives information.

1. Introduction

With the continuous increase of the management object of information archives in colleges and universities, a large number of entries have been formed, which makes it difficult to manage the archives. Nowadays, the establishment of university databases is not close to the construction of a fixed single database, but is closer to the network design of construction and centralized topology, which requires the data center to be built on the technical issues of data distributed processing and deep mining. The traditional college file management relies entirely on the manual method, the labor intensity is large, the program is relatively cumbersome and the efficiency is low.

With the continuous deepening of information, the management of college archives lags behind the needs of real development. Therefore, a scientific and rational university archives management model is constructed to realize the intelligent archives management to meet the needs of university archives management modernization. At present, the research of distributed archives management technology provides a good solution. However, when the archives management model is used to obtain the node behavior in the index, it is generally materialized, that is, it is only brought into the interactive archive information system by means of the semantic index of the materialized information center, which is limited to a single item of mechanical semantic knowledge acquisition. The research of archive index acquisition model based on dynamic fuzzy semantics is a hot topic in recent years, and the application of context awareness in pervasive computing environment is also very representative. By sensing the context information of the environment in which the node is located, the node behavior record is updated in real time, and the archive management is performed objectively and effectively on the node, so as to accurately reflect the intelligent management means of the archives, which improve the security of the network, promote the sharing of network resources, reduce the presence of malicious attacks and selfish nodes in the network, and maintain the security of the network system^[1-2].

The research content in the paper is based on the reform and innovation of college archives information management. In the grid-based university information platform, context-aware technology is applied to deal with archive index blur, and the way of the semantic conception of file index semantics is analyzed and index feature vectors is implemented under multi-dimensional

model, it aims to improve the accuracy of acquisition and make efforts to improve the level of information construction in colleges and universities^[3-4].

2. Grid University Archives Information Management System

In the network construction of colleges and universities, the regularization of the internal computer processing center and terminal platform has been formed, and the information caliber is also standardized among the platforms; the data processing ability of the isomerized platform can be realized through the grid information management system of the university; the data retrieval and retrieval of file management can be implemented by the effective mining of computer data. It is applied to the file sharing system and collaborative work central server design. The research of computer data mining technology based on big data has positive and important significance for improving the processing performance and data retrieval capability of the archives management system^[5].

The grid of the university archives information management system uses interfaces to connect scattered information resources distributed on the Internet and Web sites. The grid software links homogeneous and non-homogeneous information through various protocols and passes advanced analysis, and reduces the complexity of the technology brought about by various forms of information^[6]. The advantage of the distributed heterogeneous database unified retrieval system is to eliminate the difference in the usage of distributed heterogeneous databases, achieve cross-database unified retrieval and low resource occupancy, and not affect the smooth network bandwidth. The archive information database catalogs the meta data of various digital resources through the information grid platform. In the index query of the archives, the specific information such as the database name, the directory range, the file number, the registration date, the file key field can be retrieved, and the corresponding semantic items can be used for fuzzy search according to the context sensing. Therefore, the effective archives information can be retrieved more efficiently and accurately, and the cumbersome screening process in the archives database is avoided^[7]. The structural model of the college archives information processing in the grid, which is as shown in Figure 1.

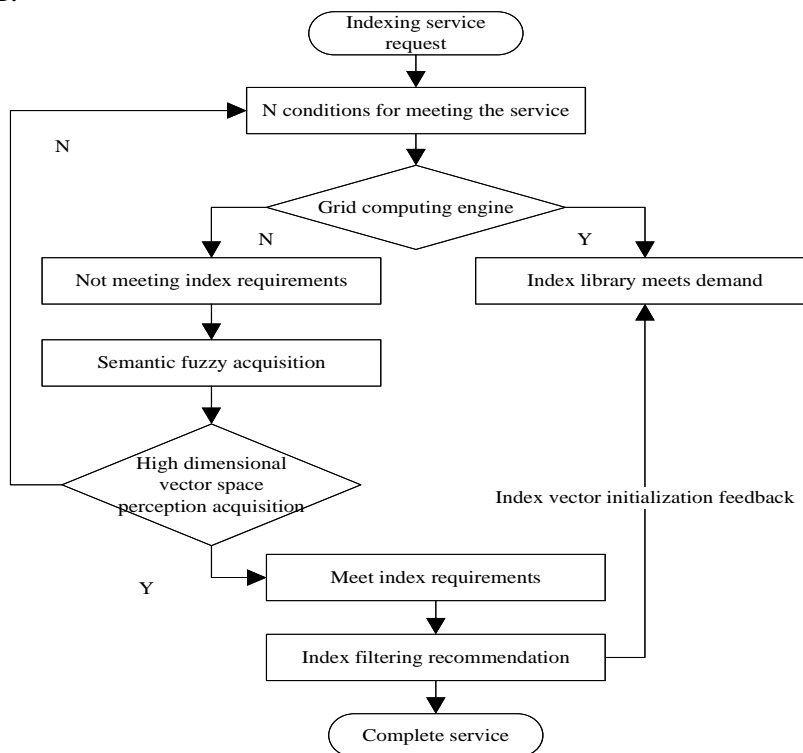


Figure 1 The structural model of college archives information processing after gridding

The gridded architecture is to study the optimization retrieval and data mining technology of

archive management information big data. First, it is necessary to establish a gridded archive management system. The archive management information is stored in the database structure model in the form of a large database format. The established archive data engine is used to drive the context-aware algorithm studied in the paper, which makes full use of the computing power and network technology advantages of the university's existing technology platform, and improves the efficiency of information management.

3. Context-aware acquisition based on college archive index

The context-aware technology of college archives management system is based on relying on archive information. According to the needs of college archives information management, semantic technology is used for information processing. The logical relationships between subjectivity and object imaginary are realized through semantic acquisition to achieve intelligent and deep access to ontology semantic index information. Depending on the context in which the context-aware technology is located and the entities it contains, any information describing its state is expressed in an implicit or explicit manner. An entity can be an archive number, an archive categorization, or a related object that interacts with an archive and an archive index, which includes colleges and college indexes.

The contextual information of λ different types in the environment is defined as follows.

$$c = (c_i, c_j, \dots, c_\lambda) \quad (1)$$

Where $ci \subset c(i \in [1, \lambda])$ represents any type of single-dimensional context information, which is specifically described as $c_i = \{c_{i_p} | p \in [1, q_i]\}$, and q_i represents the number of instance information of the context type in the dimension.

The examples of quantized context information are as follows.

$$c_i: \text{archive-number}(\text{digital}^{[1]}\text{letter}^{[2]}\text{symbol}^{[3]})$$

$$c_j: \text{archive-category}(\text{personnel}^{[1]}\text{Finance}^{[2]}\text{student status}^{[3]})$$

In the context of the university file management system, the traditional "user-project" two-dimensional scoring prediction model is extended to a multi-dimensional predictive model containing context information.

$$R: \text{archive-category} \times \text{archive-index} \times \text{archive-number} \rightarrow \text{archive-return}$$

Where, archive-category, archive-index, and archive-number are archive categories, archive indexes, and archive numbers, respectively. The structured description examples are as follows.

$$\text{archive-category} \subseteq \text{archive-level} \times \text{archive-group} \times \dots$$

$$\text{archive-index} \subseteq \text{index content} \times \text{index direction} \times \dots$$

$$\text{archive-number} \subseteq \text{number field} \times \text{QR code} \times \dots$$

Based on the above description, a method in the paper is proposed for analyzing cognitive behavior and context-aware acquisition systems of archives under the environment of context. The definition of related data is described as follows.

Definition 1: Complete project set. In the context, the provider of the project is a network service provider that contains different sets of projects and content. The project entity l of any category can be described as

$$Ent_l = \{s_r | s \in [1, n]\}$$

where N is number of projects.

A full project collection describes only the initial set of projects that are independent of a particular context type, which is the largest union of the individual project entities. Its structured description is

$$s_{vr} = [s_{vr_n} | n \in (1, m)]$$

where, $s_{vr_n} \subseteq (Ent_1 \cup \dots \cup Ent_\lambda)$, $m = Card(Ent_1 \cup \dots \cup Ent_\lambda)$ is the total number of items in the

initial project collection.

Definition 2: Multidimensional context user feature quantity. For any user U_j

$$\text{Pre}_{jis} = [\text{Pre}_{jis} | j \in (1, u)], i \in (1, k), s \in (1, n)$$

it indicates the feature quantity of the project S_s by any user U_j under the context environment C_i , where

$$S_s \in S_{vr}$$

The larger the Pre_{jis} value, the greater the probability that the item S_s meets the user U_j feature quantity.

Definition 3: Context User Feature Information Structure $\forall U_j$

$$UP_j = (U_j, S_{vr_j}, C_j, \text{Pre}S_j)$$

A feature quantity information structure describes the context user U_j , where S_{vr_j} represents the initial complete item set of the user U_j , that is, a collection of various items that the user U_j may be exposed to, C_j is the current context information, and $\text{Pre}S_j$ is the U_j feature quantity item set.

Definition 4: Formal definition of context-awareness acquisition systems. If

$$UD_{pre} = (D_{j1}, \dots, D_{j(Ent_1)}) [j \in (1, U)], Ent_1 \subseteq S_{vr}$$

is the target dimension space of the feature quantity to be predicted,

$$D_{pre} = (D_{i1}, \dots, D_{i(Ent_k)}) [i \in (1, U)], Ent_k \subseteq S_{vr}$$

is the feature quantity result dimension space, and $UD_{pre} \cap D_{pre} = \phi$.

The utility function $F_{ury}(\bullet)$ is used to calculate the feature quantity Pre_{jis} of the user under the multi-dimensional context condition. Then the context-aware acquisition system is the $d_{j1}, \dots, d_{j(Ent_1)}$, and the tuples $d_{i1}, \dots, d_{i(Ent_k)}$ with the largest feature value are found. The formal definition is as follows:

$$\forall (d_{j1}, \dots, d_{j(Ent_1)}) \subseteq UD_{pre},$$

$$(d_{i1}, \dots, d_{i(Ent_k)}) = \arg \max F_{ury}(UD_{pre}, D_{pre})$$

Through the above definition of the college archives information management system, an index acquisition method based on context-aware fuzzy recognition is established. According to the information fragment obtained by the user fuzzy, the high-dimensional analysis of the information feature vector is realized; and the calculation result with higher precision can be realized in the heterogeneous platform of the grid computing.

4. Analysis of experimental results

4.1. Test environment and rules

In the paper, according to the application test of the research results, the virtual test environment built by existing resources is established in the grid information platform of colleges and universities. In the Windows 2008 sever environment, the grid space virtualizes 10,000 node terminals, and each node provides a grid service. These nodes can provide both indexing service requests and indexing service operations, which are consistent with the actual computing environment.

The types of services provided on the nodes include CPU services, memory services, and network bandwidth services. The number of these three types of services is randomly assigned, and the service performance of each service type is different. When a node becomes a service requester,

the service is randomly sent to become a service demand end, and its resource occupation is mainly performed by the local service; and the node without the service request becomes a server for information processing. According to the setting of the virtual environment, the number of service requests is greater than the number of servers, which will become the strength of the virtual service test verification work.

The parameters in the mesh model are set according to the following constraints. The request for the archive entity that just entered the grid satisfies the initial value of 0.5. Since the update of the archive entity feature vector is event-driven, that is, after each successful acquisition of the index task, a new item is added to the entity feature vector space, the expansion coefficient of the feature library is also increased. The set value is fixed in the range of 0.69 that avoids the high expansion value to interfere with the acquisition accuracy. When the filtered entity feature database satisfies the index reference requirement, the weight of the new signature database is also weighted and averaged, which can objectively reflect the test real effect.

4.2. Test procedure

In the test process, the index acquisition of the platform entity information is first tested. In the system, 1000 nodes are randomly selected for service application, and after completing the service, they are sequentially added to 2000, 3000, ..., 5000 service applications to test the service capability and stability of the grid platform under load. After the system is in a stable state, the data collection is performed; after the entity information index acquisition task is completed, the index acquisition based on the semantic-aware behavior is performed, and the global counter records the effective number and the number of failures, and the task ends when the application service between the nodes reaches 5000 times; according to the above steps, the multi-dimensional iterative process is sequentially performed. In this test, the dimension is controlled in 5 units, and the calculation probability time is controlled within 1 second. The random iteration running test data is shown in Table 1.

Table 1. Multi-dimensional semantic perception for accurate range acquisition

| Accuracy Dimension | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|-----------------------|---------------|---------------|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Dimension1 | 0.634563 6 | 0.76546 68 | 0.678435 62 | 0.341465 3 | 0.75492 31 | 0.67533 51 | 0.75322 81 | 0.545217 1 | 0.42671 46 | 0.56724 66 |
| Dimension2 | 0.565789 9 | 0.66856 56 | 0.778937 89 | 0.613443 3 | 0.76694 32 | 0.54567 17 | 0.86576 83 | 0.452447 2 | 0.42724 44 | 0.41786 67 |
| Dimension3 | 0.866744 4 | 0.78734 49 | 0.746157 44 | 0.853562 2 | 0.78842 31 | 0.63656 71 | 0.77565 74 | 0.451611 7 | 0.47172 54 | 0.40562 42 |
| Dimension4 | 0.954465 7 | 0.76435 68 | 0.447835 67 | 0.684523 2 | 0.57892 20 | 0.89454 31 | 0.67946 46 | 0.542722 1 | 0.45762 42 | 0.41647 33 |
| Dimension5 | 0.944663 8 | 0.45834 66 | 0.525611 33 | 0.563445 2 | 0.75832 01 | 0.63621 28 | 0.67533 82 | 0.457828 1 | 0.46161 18 | 0.50541 23 |

The semantic-aware index acquisition obtained by the content of table 1 can be referenced to compare the entity index, and the accuracy range is divided according to the quantity level as shown in figure 2.

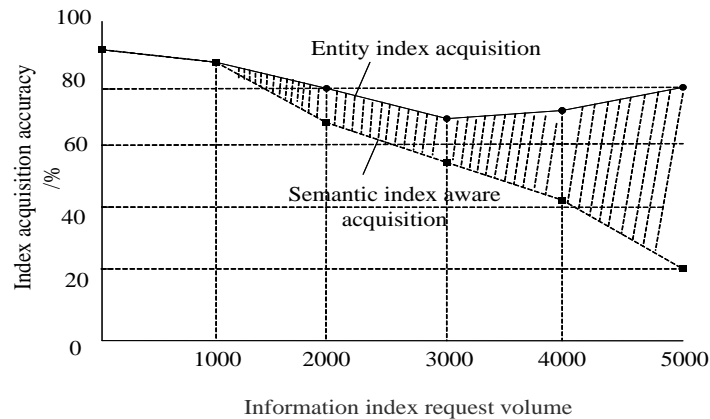


Figure 2. Semantic perception acquisition accuracy comparison reference

The results of this environment are shown in Figure 2. For the semantic acquisition of 5000 units, the accuracy is over 3,000 units over 57%, and the test accuracy of 1000 units is completely consistent with the entity index. The acquisition accuracy of 1000 units can be corrected by the second iteration in the index acquisition of 40%~57%. The index acquisition with the accuracy lower than 40% is regarded as the number of failures, and the accuracy of this part of the data is improved, which needs to be solved in technical research such as more in-depth algorithm optimization in the future. The experiment proves that the fuzzy perception technology studied in the paper basically meets the application conditions and can meet the requirements of college archives information management.

5. Conclusions

In the process of college archives information construction, the research content of the paper makes full use of the existing resources of colleges and universities to combine the grid computing platform suitable for campus development, and adopts context-aware technology to solve the technical problems of accurate indexing of college archives. Aiming at the indexing requirements of fuzzy conceptualization, the reference effect of materialized index feature vectors in multi-dimensional initial sets is studied. The proposed optimization definition principle is beneficial to the further study of college archives information processing.

References

- [1] Reputation Measurement and Malicious Feedback Rating Prevention in Web Service Recommendation Systems. Wang Shangguang, Zheng Zibin, Wu Zhengping, et al. IEEE Trans. Services Computing. 2015
- [2] Integrating Context Similarity with Sparse Linear Recommendation Model. Yong Z, Bamshad M, Robin B. International Conference on Web Information Systems Engineering. 2015
- [3] Context-Aware Restaurant Recommendation for Natural Language Queries: A Formative User Study in the Automotive Domain. Philipp F, Cornelius S, Andreas N. Conference of the International Speech Communication Association. 2016
- [4] Context-Aware Recommendation Via Graph-Based Contextual Modeling and Post filtering. Wu H, Yue K, Liu X, et al. International Journal of Distributed Sensor Networks. 2015
- [5] Geo-SAGE: A Geographical Sparse Additive Generative Model for Spatial Item Recommendation. Wang W, Yin H, Chen L, et al. Data Mining and Knowledge Discovery. 2015
- [6] A grid portal with robot certificates for bioinformatics phylogenetic analyses. Barbera, R., Andronico, G, Donvito, G, Falzone, A., Keijser, J.J., La Rocca, G, Milanese, L., Maggi, G.P., Vicario, S. Concurrency Computation Practice and Experience. 2011
- [7] Grid Cert Lib: A Single Sign-on Solution for Grid Web Applications and Portals. Murri, Riccardo, Kunszt, Peter Z., Maffioletti, Sergio, Tschopp, Valery. Journal of Grid Computing. 2011