ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE

Dynamic semantic description and search methods for heterogeneous learning resources

Xiaocong LAI, Ying PAN*, and Xueling JIANG

Guangxi Key Lab of Human-machine Interaction and Intelligent Decision, Nanning Normal University, Nanning 530001, People's Republic of China

Abstract. Learning resources are massive, heterogeneous, and constantly changing. How to find the required resources quickly and accurately has become a very challenging work in the management and sharing of learning resources. According to the characteristics of learning resources, this paper proposes a progressive learning resource description model, which can describe dynamic heterogeneous resource information on a fine-grained level by using information extraction technology, then a semantic annotation algorithm is defined to calculate the semantic of learning resource and add these semantic to the description model. Moreover, a semantic search method is proposed to find the required resources, which calculate the content with the highest similarity to the user query, and then return the results in descending order of similarity. The simulation results show that the method is feasible and effective.

Key words: heterogeneous data; learning resources; semantic description; semantic search.

1. INTRODUCTION

Learning resources are an important part of educational activities, and an important factor to improve the quality and effectiveness of teaching. With the rise of digital education and the arrival of the era of big data, learning resources are growing explosively, which needs to be effectively managed. Learning resources have the following characteristics: a) multi-source and heterogeneous; Massive digital learning resources involve a variety of heterogeneous data resources, which can be called heterogeneous data from multiple sources. For example, there are many formats for data, such as text format and XML format, and the data structure includes structured data, semi-structured data, and unstructured data. b) rich semantics; Learning resources contain rich semantics, and there are many semantic relationships among them, such as precursor relationship, inclusion relationship, equivalence relationship, etc. c) dynamic change; The evolution of the content of learning resources makes resources in constant dynamic change. Therefore, the management and sharing of learning resources must meet the real-time needs, and the updated content can be searched in time.

How to find the required learning resources quickly and accurately has become a research hotspot. Semantic search research has introduced various representations of semantic information, such as semantic tags, ontology representation in the Semantic Web, and semantic information retrieval [1]. However, existing semantic search methods are difficult to describe and search dynamic heterogeneous learning resources in a timely manner, and it is difficult to find appropriate learning resources accurately and quickly according to learning needs.

To solve these problems, this paper proposes a progressive semantic description method, which can effectively and clearly describe the dynamic heterogeneous resource information. On this basis, a semantic search method is proposed, which can quickly and accurately find the required resources.

The remainder of this paper is structured as follows: Section 2 introduces the research status of dynamic semantic description and search. Section 3 introduces the design and implementation of dynamic semantic description of learning resources. Section 4 introduces the design and implementation of semantic search methods. Section 5 makes an experimental analysis of the proposed methods. The last section gives conclusions and future work.

2. RELATED WORK

2.1. Description and search of learning resources

For many years, scholars have devoted themselves to the study of learning resource description methods. IMS Common Cartridge [2] used files to organize and manage the content and structure of resources. LOM [3] (Learning Object Metadata) specified a conceptual model that defined the descriptive, structural, and semantic features for a learning object. SCORM [4] (Sharable Content Object Reference Model) proposed a resource description and encapsulation model to describe micro granularity resources, further refine the shareable granularity of learning resources, and provide a smaller resource sharing method than file level. Some scholars consider the evolution of learning resources in their resource descriptions. Yu *et al.* [5] presented an organizational model for organizing learning resources. By using time dimension and knowledge ontology, Learning Cell can support the dynamic evolution of learning resources, and describe the internal structure and external relations of learning resources more flexibly. Chen *et al.* [6] proposed a dynamic

---
*e-mail: panying@gxtc.edu.cn

learning resource model with context. The model has context information and flexible content, which allows content to be dynamically adjusted in different contexts, thereby providing learners with the most appropriate resources.

In order to describe the rich semantics of learning resources and enhance the retrieval, reuse and combination of resources, semantic web technology, especially ontology technology, has been widely used in the field of education [7, 8]. Sein-Echaluce *et al.* [9] proposed a semantic search system based on Web 3.0. Subject resources are organized and classified by ontology, and then the searchability of resources is improved by ontology reasoning, so that students can better acquire the knowledge in the subject. Palombi *et al.* [10] constructed an ontology-based learning management system, which used RDF (Resource Description Framework) to describe the important vocabulary of educational content, students' activity traces and other information, and then, SPARQL query language is used to support data query and analysis of the specific needs of users. In order to discover, manage, and incorporate open education resources, Koutsomitropoulos *et al.* [11] introduced the concept of learning object ontology repositories to enhance and maintain the metadata of educational resources, and then annotate and discover educational resources by reference semantic thesauri. Cerón-Figueroa *et al.* [12] introduced a new model of pattern classification and its application to align instances from different ontologies, which are related to the content of e-learning education in the knowledge society, and the experimental data here involves the OWL (Web ontology language) ontology language format and the LOM learning object metadata format. Vagliano *et al.* [13] proposed a platform which included a huge amount of heterogeneous educational resources, such as documents, videos, and social media posts. The platform represents users and documents using thesaurus or ontology, and recommends educational resources based on users' search history. In addition, with the increasing abundance of learning resources, the dimension of semantic space increases sharply. Using dimension reduction techniques to select representative semantics will speed up the query speed [14].

In summary, the application of ontology in the educational environment mainly focuses on resource description and resource retrieval. However, ontology construction is a complicated process. It is difficult or even impossible to create and use ontology for a specific domain. The construction process of ontology in the field of education also faces such challenges and difficulties [15, 16].

2.2. Dynamic semantic search

In the dynamic search scenario, the system usually improves an initial query (such as modifying and expanding it with semantic information or constraints) to make the initial request more precise. The problem of computing the rewriting of an extended query is studied [17], for avoiding recomputation, the initial query is 'decomposed' into its atoms and then each atom is processed incrementally. In order to develop KBC (Knowledge Base Construction) system, Shin *et al.* [18] proposed an incremental reasoning method based on sampling and variational techniques, which can incrementally generate semantic

reasoning results of KBC system. Kang *et al.* [19] proposed an incremental optimization method based on semantic rule guidance for the periodic query scenario of the data warehouse. By extending the query syntax, users can describe the repeated cycle and increment table of the query, and the optimized incremental query plan can be executed on the mainstream distributed computing framework (such as Map Reduce). However, for the large-scale database with high-frequency updates, the query performance of the current incremental processing method still has more room for improvement [20].

In addition, many scholars have integrated the idea of dataspace [21] into the research of dynamic semantic search. The core idea of data space is to manage data by PAYG (pay-as-you-go) mode. The characteristics of this mode are it can provide simple functions (such as keyword query) without or only at a very low cost of early construction, and then provide gradually enhanced data management functions (such as semantic query) according to the cost of investment. Salles *et al.* [22] proposed a PAYG query processing strategy: through matching, transformation, and merging, it can extend the semantics of the source query to get a new query, so as to improve the query effect. However, it is difficult to describe complex semantics, and the cost of query rewriting is high. Mahmoud *et al.* [23] proposed the schema clustering and retrieval of multi-domain systems, which implements PAYG mode data management by dynamic and progressive data integration of various coexisting heterogeneous data. Belhajjame *et al.* [24] studied how to gradually select the mapping that satisfies the query requirements through user feedback. By dynamically expanding the query range with feedback information, Yuan *et al.* [25] proposed a retail product search and recommendation method that embeds weighted TF-IDF in keywords. For the ontology-based data access/query system, Sequeda *et al.* [26] proposed the ontology construction and mapping method based on PAYG mode. Curry *et al.* [27] discussed PAYG data management in an intelligent environment, and the entity-centric query service of Real-time Linked Dataspace was proposed to meet the interactive query delay requirements, but it cannot achieve a large-scale implementation and the application cost is high. In addition, Zhou *et al.* [28] combined RDF triplet storage with OWL 2 reasoner and proposed an OWL query method based on PAYG mode, which can handle the extension of any OWL ontology. In order to provide scalable PAYG query answering, Zhou *et al.* [29] presented a hybrid approach to query answering over OWL 2 ontologies that combined a Datalog reasoner, that is, the approach delegated most of the computation to the Datalog reasoner and used the expensive OWL 2 reasoning only when queries need to be fully answered. The above research implements OWL query based on PAYG mode, but the query effect depends on the maturity of OWL 2 reasoner.

There has been some research close to our work in heterogeneous data fusion and integration. Zhang *et al.* [30] discussed the challenges of dealing with multi-source heterogeneous data fusion and introduces the deep learning methods for heterogeneous data fusion. Hu *et al.* [31] proposed a virtual data space model describing multi-source heterogeneous data resources, presented the automatic construction process of the model, and

designed a dynamic evolution algorithm to track the life cycle of data resources in real time. Kiran *et al.* [32] discussed the limitations of the existing IE (Information extraction) techniques due to the heterogeneity and unstructured big data and proposed a potential solution to improve the unstructured big data IE system, which can increase the efficiency and effectiveness of the data analysis process. Sakouhi *et al.* [33] proposed a new semantic trajectory annotation method, which integrated the original mobile data, geographic information, and social media. The results showed that the method improved the quality of annotation words by integrating two semantic sources. To improve search efficiency for detecting singleton review spammers, Isa *et al.* [34] presented a new scalable windowing approach for pairwise-similarity search, and an in-depth search strategy that established a relationship between twitter status and user engagement after a tweet has been posted. The results showed the interesting symmetry properties in terms of similarity distribution and duration.

To sum up, the main problems existing in the description and query research of learning resources are as follows: a) The methods of resource description lack the dynamic expansibility and cannot describe the heterogeneous and dynamic learning resources in time. b) The current education semantic description methods have high cost and high time complexity in query processing. In the recent years, the research of heterogeneous and dynamic data integration and management has made great progress. We plan to introduce the idea of these achievements to solve the above problems of learning resources description and query.

3. DYNAMIC SEMANTIC DESCRIPTION FOR LEARNING RESOURCES

3.1. The description model of learning resources

According to the characteristics of learning resources, we propose a learning resources description model LRD (Learning Resources Description) which is defined as follows:

Definition 1. Learning resources are described by a directed graph LRD $G := (N, E)$, where N is a set of nodes $N_1, \ldots, N_k$, each node N_i is a set of attribute-value pairs, each value can be a bag of words or text content. N_i called blank node if $N_i = \Phi$, where the set of attributes of each node may be different, and even there exist the nodes without attribute-value pairs (blank nodes) which are used to describe incomplete information. E is a set of labeled directed edges (N_i, N_j, L), where $N_i, N_j \in N$, $i \neq j$ and L is a label which can be a *null* value. Edge can be used to describe any kind of relationships between nodes, and L is a *null* value if the edge has not an explicit label. Specifically, when there is not any connection between nodes, $E = \Phi$.

Learning resources usually include structured, semi-structured, and unstructured data such as audio, video, pictures, question banks, and test papers. With information extraction technology, LRD can describe the different formats of learning resources.

Example 1. LRD describes the question bank documents (see Fig. 1).

From Fig. 1, it can be seen that the nodes of LRD represent the file directory "question banks", the files "C++ question-bank.doc" and "Java question-bank.pdf". The node "question banks" is a set of attribute value pairs {(type, "folder"), (size, 69835), (createddate, 12/09/2018 13:04), (updateddate, 25/10/2019 18:45) . . . }. The node "java question-bank.pdf" is a set of attribute value pairs {(type, "PDFfile"), (size, 25420), (createddate, 23/09/2018 10:34), (updateddate, 25/10/2018 11:42), (author, "MingLi"), (version, . . .) . . . }. LRD can also describe the fine-grained contents of the documents, such as the choice questions and their options. Furthermore, Edges represent the associations among the files and their contents. For example, the file "C++ question-bank.doc" is in the "question banks" directory, which can be expressed as: (question banks, C++ question-bank.doc, has file); the association of question and the option B can be expressed as (question NO 1 . . . , B . . . , has option).
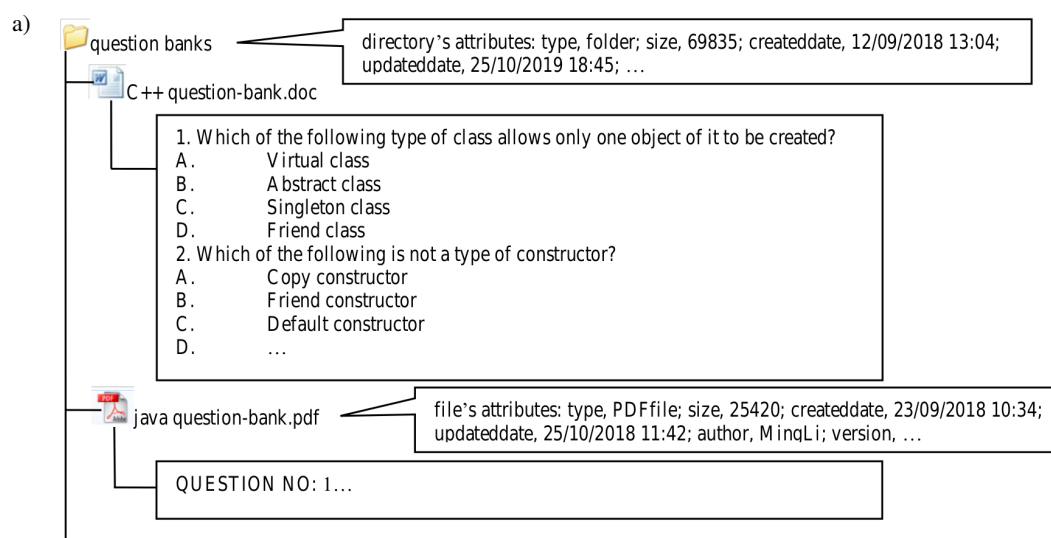

Fig. 1. LRD description of the question bank documents: a) question bank documents

Bull. Pol. Acad. Sci. Tech. Sci., vol. 70, no. 3, p. e139434, 2022

3

Xiaocong Lai, Ying Pan, and Xueling Jiang
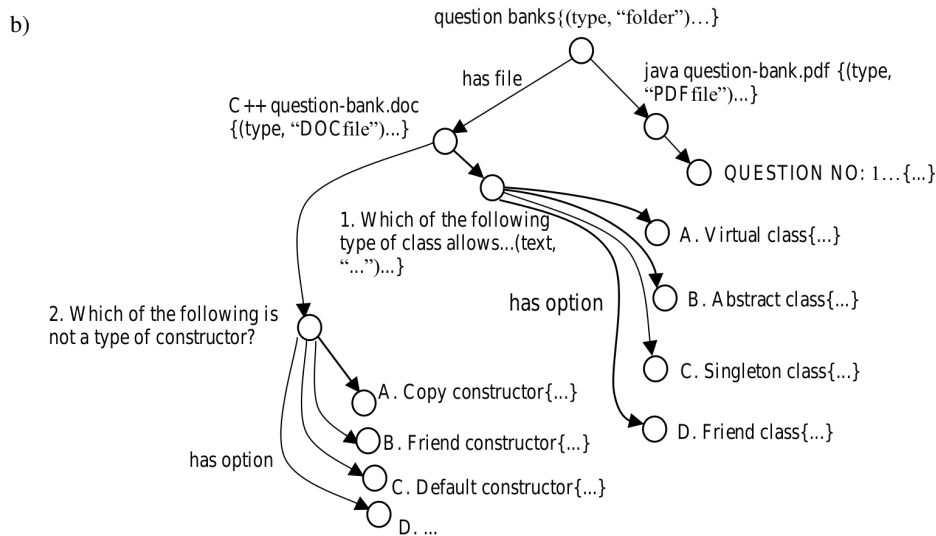
b)

Fig. 1. LRD description of the question bank documents: a) question bank documents, b) LRD

3.2. Dynamic semantic description method

The description of learning resources by LRD is a dynamic process. Semantic information about resources is added to LRD through repeated optimization. The specific idea is shown in Fig. 2.
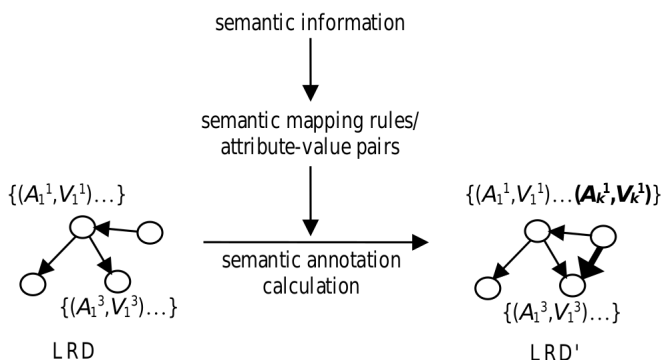

Fig. 2. A dynamic process of semantic description

The semantic information can be obtained through various methods such as machine learning and text content mining. We mine the attributes and their values of this information to generate attribute value pairs in the form of N{(attribute, value)...}. Moreover, we analyze the semantics of this information, such as synonymy, antisense, global, local, etc., and generate semantic mapping rules from simple to complex asymptotically. The semantic mapping rules are as follows: C same-as D (reflects the synonymous relationship between C and D); C different-from D (reflects the antisense relationship between C and D); C is–a D (reflects C as a member of D); C part-of D (reflects C as a part of D); C is-instance D (reflects C as an example of D).

We define the semantic annotation algorithm (see algorithm 1) to calculate and return the content of the semantic annotation (attribute-value pairs or edges of nodes). The algorithm determines whether the semantic information is related to the node or the edge in LRD. If it is related to the node, the content of the semantic annotation is calculated and returned in the form of the attribute value pair of the node. If it is related to the edge, the content of the semantic annotation is calculated and returned in the form of the edge between nodes.

We add the above semantic annotation to LRD to generate the LRD' model with semantic information (as shown in the bold arrow and attribute value pair $(A_k^1 V_k^1)$). The adding method is as follows: The semantic association between nodes is represented by edges, and the semantics of node attributes are added to attribute value pairs in an extended way.

Algorithm 1. Semantic annotation

Input: LRD, semantic mapping rules {C is-a D, ... },
semantic information attribute-value pairs M{(attribute, value) ... };
Output: LRD'.

1 for $i = 1$ to k do
2   Compare C, D, M and N_i
3   if $C = N_i$ or $D = N_i$ then
       rewrite LRD to LRD' by adding an edge with label
       "is-a" between nodes N_i and C/D
4   if $M = N_i$ then
       rewrite LRD to LRD' by adding the attribute-value
       pairs of M to N_i in the form of N_i {(attribute, value) ... }
5 end do
6 return LRD'

---

4. SEMANTIC SEARCH METHOD

The way of semantic annotation does not change the structure of LRD, so our query method can support semantic query on LRD' without special extension, that is, this query method is universal to LRD or LRD'.

We use the similarity function to describe the similarity of two learning resource documents.

4

Bull. Pol. Acad. Sci. Tech. Sci., vol. 70, no. 3, p. e139434, 2022

Definition 2. The semantic similarity function of nodes $C\text{sim}(N_i, N_j)$ and its threshold t describe the similarity between nodes on LRD, where $t \in (0, 1]$, $N_i, N_j \in N$.

The similarity function $C\text{sim}(N_i, N_j)$ can be calculated by traditional IR (Information Retrieval) technology (such as TF (Term Frequency)/IDF (Inverse Document Frequency) and cosine similarity method). For example, the similarity of two learning resources can be calculated as follows: extract two groups of words (such as the name, author, and file type of learning resources) from two learning resources nodes respectively, and then use the cosine similarity method to calculate the similarity of these two groups of words.

We mainly consider attribute similarity and attribute-value similarity of nodes, and define the calculation of $C\text{sim}(N_i, N_j)$ in equation (1) as follows:

$$C\text{sim}(N_i, N_j) = W_1 \times A\text{sim}(A, B) + W_2 \times V\text{sim}(V_A, V_B)$$
$$= W_1 \frac{\sum_{k=1}^{n} a_k b_k}{\sqrt{\sum_{k=1}^{n} a_k^2} \sqrt{\sum_{k=1}^{n} b_k^2}} + W_2 \frac{\sum_{k=1}^{n} Va_k Vb_k}{\sqrt{\sum_{k=1}^{n} Va_k^2} \sqrt{\sum_{k=1}^{n} Vb_k^2}}, \quad (1)$$

where, W_1, W_2 are the weights, and $W_1 + W_2 = 1$. A and B are the attribute sets of nodes N_i and N_j, respectively, in the form of $\{a_1, a_2, \cdots\}$, $\{b_1, b_2, \cdots\}$, V_A, V_B are the attribute value sets of A and B, respectively, in the form of $\{Va_1, Va_2, \cdots\}$, $\{Vb_1, Vb_2, \cdots\}$. $A\text{sim}(A, B)$ is the attribute similarity function, $V\text{sim}(V_A, V_B)$ is the attribute value similarity function, they are calculated by the cosine similarity method.

Definition 3. The semantic similarity function of edges $R\text{sim}(R_i, R_j)$ and its threshold t describe the similarity between edges on LRD, where $t \in (0, 1]$, $(R_i, R_j) \in E$.

We mainly consider the similarity of graph structure and the label similarity of edges, and define the calculation of $R\text{sim}(R_i, R_j)$ in equation (2) as follows:

$$R\text{sim}(R_i, R_j) = 1 - \sqrt{\frac{\text{ged}(R_i, R_j)}{\max\{M_{Ri}, M_{Rj}\} + E_{Ri} + E_{Rj}}}, \quad (2)$$

where, the correlation R_i contains M_{Ri} nodes and E_{Ri} edges, the correlation R_j contains M_{Rj} nodes and E_{Rj} edges, and $\text{ged}(R_i, R_j)$ is the graph editing distance function, which is used to calculate the minimum cost of the complete editing path from the source graph to the target graph.

The calculation method of $\text{ged}(R_i, R_j)$ is as follows. R_i and R_j can be regarded as a graph with related nodes and edges respectively. The editing distance between two graphs refers to the sum of the costs of the minimum editing operations required to transform one graph into another graph. The graph editing operations are specifically divided into the following six basic operations: a) insert an isolated label node; b) delete an isolated node; c) replace the label of a node; d) insert a label edge to two unconnected nodes; e) delete a label edge; f) replace the label of one edge in the figure. Assuming that the cost of all editing operations is 1, the graph editing distance is calculated and

compared with the given threshold value to judge the similarity between the two graphs. The smaller the graph editing distance is, the higher the similarity will be.

We implement the semantic keyword query on LRD' through the following steps: When the user queries, we analyze the user's query requirements based on natural language analysis and keyword extraction methods, then calculate the node set and information to be compared on LRD', and then use the semantic similarity algorithm (see algorithm 2) to calculate the node with the highest similarity to the user query, and then return the results to the user in descending order of similarity.

Algorithm 2. Semantic keyword query

Input: keyword query k; threshold t.

Output: result set NoteRes (attribute-value pairs of nodes) or EdgeRes (edges).

1 for each node $n \in N$ do
2   compute $C\text{sim}(k, n)$
3   if $C\text{sim}(k, n) \geq t$ then NoteRes := NoteRes $\cup n$
4 end for
5 for each edge $e \in E$ do
6   compute $R\text{sim}(k, e)$
7   if $R\text{sim}(k, e) \geq t$ then EdgeRes := EdgeRes $\cup e$
8 end for
9 return NoteRes/EdgeRes that are ranked by a descending order
   of their similarity

---

5. EVALUATION

In this section, the simulation experiments have been performed to evaluate the performance of semantic description and search approach mentioned in Section 3 and Section 4.

The computer used for the experiments is Intel(R) Pentium(R) CPU G4560 @ 3.50GHz with 4GB of RAM. By extracting some learning resources, we constructed LRD which contained 2160 nodes and 2165 edges, then we constructed LRD' by algorithm 1 which contained 2436 nodes and 2439 edges. We defined six keyword queries that involve nodes and edges. We set the threshold $t = 0.7$, this value is generally accepted in similarity calculations [35].

The six keyword queries on LRD/LRD' and their response times are shown in Table 1. The response times are obtained on a warm cache. From the experimental result, we can see that most queries response times on LRD are less than those on LRD'. This is because LRD' with semantic information has more sides and nodes than LRD, so it takes more time to query on LRD'. Taken as a whole, the gap of response times between LRD and LRD' is very limited. Thus, we can draw the conclusion that our approaches do not add much overhead for semantic annotation.

We evaluated the results by using the IR metrics recall, precision, and F-measure, which are the important basis for measuring the effectiveness of information retrieval. The relevant concepts and calculations are as follows:

Recall R is the ratio of the amount of relevant information queried to the total amount of relevant information in the retrieval system:

Bull. Pol. Acad. Sci. Tech. Sci., vol. 70, no. 3, p. e139434, 2022

5

Xiaocong Lai, Ying Pan, and Xueling Jiang

R = (Number of related information queried/total number of related information in the system) $\times 100\%$

Precision P refers to the ratio of the amount of related information queried to the amount of all information queried:

P = (Number of related information queried/total number of query results) $\times 100\%$

F-measure comprehensively considers the effect of recall and precision, and its calculation is as follows (3):

$$F_{\beta} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}. \tag{3}$$

Here β is the parameter, when $\beta = 1$, F-measure is obtained according to equation (4):

$$F = \frac{2 \times PR}{P + R}. \tag{4}$$

Figure 3 shows the recall and precision of queries on LRD/LRD'. We can see that the recall or precision of most queries on LRD' are improved in different degree. For example, both the recall and the precision of Q1 on LRD' are improved greatly. However, both the recall and the precision of Q3 on LRD' are barely improved because the semantic description of a PDF file is less in LRD', and the accuracy of semantic query is not high. In general, the more semantic descriptions, the better the query effect. In summary, our semantic description and search method can improve the quality of query results, and the degree of improvement depends on the amount of semantic information.
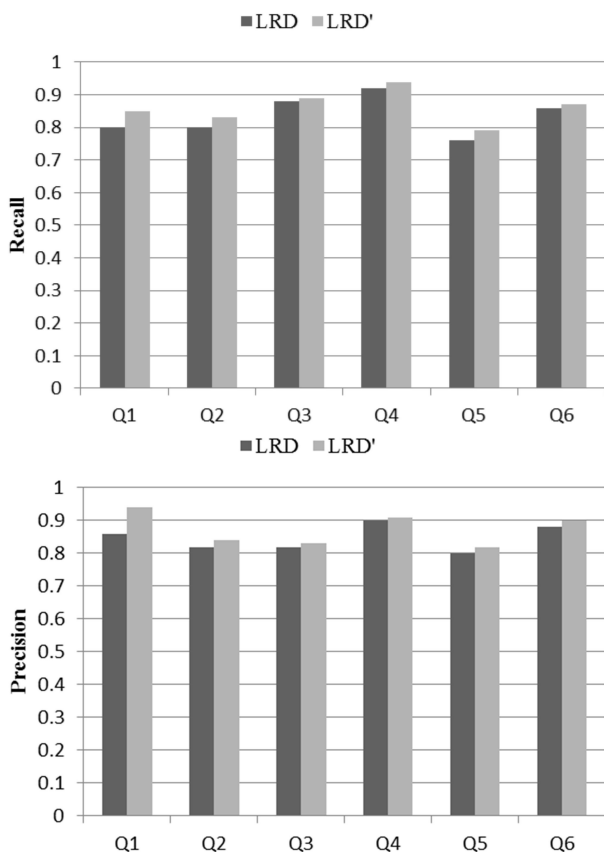
F-measure combines the results of recall and precision, which are closely related and mutually restricted. Since the query method cannot be perfect, when we need to retrieve more related documents, that is, the larger the R value, the accurate result P of retrieval will be affected. Similarly, when we want to get a more accurate result P, we will require a stricter "retrieval strategy", which will make some relevant documents not be retrieved, and R will also be affected. Therefore, this paper chooses an appropriate degree F for the "retrieval strategy" according to the needs, which cannot be too strict or too loose, and seeks a balance between R and P. Figure 4 shows that, overall, the value of F-measure on LRD' is higher than that on LRD, especially the value of F corresponding to Q1, Q2 and Q5 is significantly higher, while the F value of Q3, Q4 and Q6 is slightly higher. In addition, because the F value of LRD' fluctuates between 0.8 and 0.95, while the F value of LRD fluctuates between 0.75 and 0.95, so the curve of the value of F on LRD' is relatively stable than that on LRD. It can be seen that the query method in this paper improves the query quality, and the query strategy is ideal.

In addition, we evaluated the scalability of the proposed method through the data sets of different sizes. Firstly, by adding nodes and edges to LRD', we constructed 4 data sets: S1, S2, S3 and S4, where S1 contained 3570 nodes and 4218 edges, S2 contained 5013 nodes and 5901 edges, S3 contained 7032 nodes and 8359 edges, and S4 contained 10130 nodes and 12947 edges. Then Q1~Q6 queries were executed in the above different data sets, and the average query response time was calculated. The results are shown in Fig. 5.
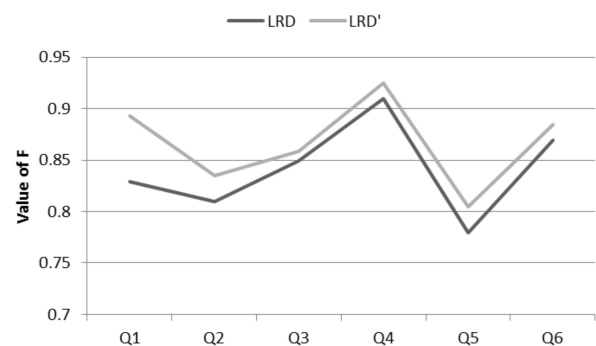

Fig. 4. The results of F-measure on LRD/LRD'

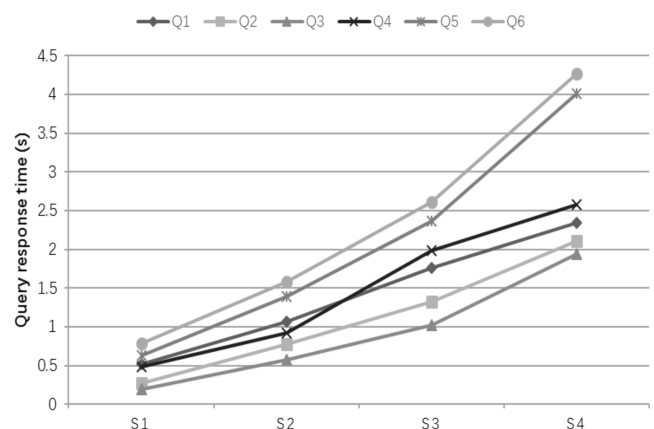Fig. 3. The recall and precision of queries on LRD/LRD'

Fig. 5. Performance of our query method on different data set scales

It can be seen from Fig. 5 that the response time of all queries increases when the size of the data set increases. The average query response time of Q1~Q4, in which query involving nodes or attributes, increases slightly, while the average query response time of Q5 and Q6, which query involving edges, increase significantly, because Q5 and Q6 spend more time on the selection and calculation of query paths. In general, the response time of all queries can still maintain linear growth, so our method has better scalability and extensibility.

6. CONCLUSIONS

In this paper, a semantic description method is presented to dynamically describe learning resources, it can describe the heterogeneous and fine-grained contents of learning resources. A semantic search method is also presented to improve the quality and accuracy of keyword-based query. The simulation results show that our methods are feasible and effective. In the future, we will plan to study two aspects: a) How to implement more complex semantic query for learning resources; b) How to describe and query learning resources with weak semantic relationship. In the following work, we plan to introduce more advanced reasoning engine and machine learning method to improve the semantic query of learning resources.

APPENDIX

Table 1
The response times of keyword queries on LRD/LRD'

Query ID	Expression	Response time on LRD (s)	Response time on LRD'(s)
Q1	question bank	0.39	0.43
Q2	java	0.18	0.18
Q3	PDFfile	0.12	0.12
Q4	version	0.28	0.29
Q5	is–a	0.23	0.24
Q6	hasfile	0.36	0.36

REFERENCES

[1] K. Mahmood, M. Rahmah, M. Ahmed, M.A. Raza, "Semantic Information Retrieval Systems Costing in Big Data Environment," in *Recent Advances on Soft Comput., and Data Mining SCDM*), 2020, pp. 192–201, doi: 10.1007/978-3-030-36056-6_19.

[2] IMS, "IMS Common Caartridge Profile v1.0," [Online]. Available: http://www.imsglobal.org/cc/ccv1p0/imscc_profilev1p0. html#0_pgfId-1752945 [Accessed: 1. Oct. 2008].

[3] LTSC (Learning Technology Standards Committee). "Draft standard for learning object metadata," USA IEEE standard 1484.12.1, July 15, 2002.

[4] S. Riley, L.K. Soh, and A. Samal, "On the Importance of Metadata and Learner Interaction Data in SCORM 2.0," LETSI SCORM 2.0 White papers, 2008. [Online]. Available: https://21w98o3yqgi738kmv7xrf9lj-wpengine.netdna-ssl.com/wp-content/assets/LETSI-white-papers/LETSI%20White%20 Papers%20-%20Riley%20-%20   Importance%20of%20Meta data.pdf

[5] S. Yu, X. Yang, G. Cheng, and M. Wang, "From learning object to learning cell: A resource organization model for ubiquitous learning," *J. Educ. Technol. Soc.*, vol. 18, no. 2, pp. 206–224, 2015.

[6] M. Chen, S.Q. Yu, and F.K. Chiang, "A dynamic ubiquitous learning resource model with context and its effects on ubiquitous learning," *Interact. Learn. Environ.*, vol. 25, no. 1, pp. 127–141, 2016, doi: 10.1080/10494820.2016.1143846.

[7] J. Jensen, "A systematic literature review of the use of semantic web technologies in formal education," *Br. J. Educ. Technol.*, vol. 50, no. 2, pp. 505–517, 2019, doi: 10.1111/bjet.12570.

[8] R. Welch, N. Taylor, and M. Gard, "Environmental attunement in the health and physical education canon: emplaced connection to embodiment, community and 'nature'," *Sport Educ. Soc.*, vol. 26, no. 4, pp. 349–362, 2021, doi: 10.1080/13573322.2021.1890572.

[9] M. L. Sein-Echaluce, Á. Fidalgo-Blanco, and J. Esteban-Escaño, "Technological ecosystems and ontologies for an educational model based on Web 3.0," *Univ. Access Inf. Soc.*, vol. 18, no. 3, pp. 645–658, 2019, doi: 10.1007/s10209-019-00684-9.

[10] O. Palombi *et al.*, "OntoSIDES: Ontology-based student progress monitoring on the national evaluation system of French Medical Schools," *Artif. Intell. Med.*, vol. 96, pp. 59–67, 2019, doi: 10.1016/j.artmed.2019.03.006.

[11] D. A. Koutsomitropoulos and G. D. Solomou, "A learning object ontology repository to support annotation and discovery of educational resources using semantic thesauri," *IFLA J.*, vol. 44, no. 1, pp. 4–22, 2018, doi: 10.1177/0340035217737559.

[12] S. Cerón-Figueroa *et al.*, "Instance-based ontology matching for e-learning material using an associative pattern classifier," *Comput. Hum. Behav.*, vol. 69, pp. 218–225, 2016, doi: 10.1016/ j.chb.2016.12.039.

[13] I. Vagliano and S. Nazir, "Recommending multimedia educational resources on the MOVING platform," *Proceedings of the 8th Workshop on Bibliometric-enhanced Information Retrieval*, 2019, pp. 148–158.

[14] C. Kumar and R. Santhosh, "Effective information retrieval and feature minimization technique for semantic web data," *Comput. Electr. Eng.*, vol. 81, pp. 1–14, 2020, doi: 10.1016/j.compeleceng. 2019.106518.

[15] G. Kaya and A. Altun, "Educational ontology development," in *Encyclopedia of Information Scienceand Technology*, M. Khosrow-Pour, D.B.A. (Ed.), Hershey: IGI Global, 2018, pp. 26–37, doi: 10.4018/978-1-5225-2255-3.ch124.

[16] B. K. Daniel, "Big Data and data science: A critical review of issues for educational research," *Br. J. Educ. Technol.*, vol. 50, no. 1, pp. 101–113, 2019, doi: 10.1111/bjet.12595.

[17] T. Venetis, G. Stoilos and G. Stamou, "Query extensions and incremental query rewriting for OWL 2 QL ontologies," *J. Data Semant.*, vol. 3, no. 1, pp. 1–23, 2014, doi: 10.1007/s13740-012-0017-6.

[18] J. Shin, S. Wu, F. Wang, Ch. De Sa, C. Zhang, and Ch. Ré, "Incremental knowledge base construction using deepdive," *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, vol. 8, no. 11, pp. 1310–1321, 2015, doi: 10.14778/2809974.2809991.

[19] Y. L. Kang, F. Li, and L. Wang, "Incremental Optimization Method for Periodic Query in Data Warehouse," *J. Software*, vol. 28, no. 8, pp. 2126–2147, 2017, doi: 10.13328/j.cnki.jos.005107.

[20] Z. Tang, M. He, L. Yang, and Z. Fu, "i2Graph: An Incremental Iterative Computation Model for Large Scale Dynamic Graphs," *2019 IEEE International Conference on Parallel & Distributed Processing with Applications*, 2019, vol. 1, pp. 654–659, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00099.

[21] M. Franklin, A. Halevy, and D. Maier, "From databases to dataspaces: a new abstraction for information management," *ACM SIGMOD Rec.*, vol. 34, no. 4, pp. 27–33, 2005, doi: 10.1145/1107499.1107502.

[22] M. A. Salles, J.-P Dittrich, S. K. Karakashian, O. R. Girard, L. Blunschi, "iTrails: Pay-as-you-go Information Integration in Dataspaces," *Proceedings of the 33rd International Conference on Very Large Data Bases*, 2007, pp. 663–674.

[23] H. A. Mahmoud and A. Aboulnaga, "Schema clustering and retrieval for multi-domain pay-as-you-go data integration systems," *Proceedings of ACM SIGMOD International Conference on Management of Data*, 2010, pp. 411–422, doi: 10.1145/1807167.1807213.

[24] K. Belhajjame *et al.*, "Incrementally improving dataspaces based on user feedback," *Inf. Syst.*, vol. 38, no. 5, pp. 656–687, 2013, doi: 10.1016/j.is.2013.01.006.

[25] T. T. Yuan and Z. Zhang, "Merchandise recommendation for retail events with word embedding weighted Tf-idf and dynamic query expansion," *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 1347–1348, doi: 10.1145/3209978.3210202.

[26] J. F. Sequeda, W. J. Briggs, D. P. Miranker, and W. P. Heideman, "A Pay-as-you-go Methodology to Design and Build Enterprise Knowledge Graphs from Relational Databases," *International Semantic Web Conference*, 2019, pp. 526–545, doi: 10.1007/978-3-030-30796-7_32.

[27] E. Curry, W. Derguech, S. Hasan, Ch. Kouroupetroglou, U. ul Hassan, "A real-time linked dataspace for the internet of things: enabling "pay-as-you-go" data management in smart environments," *Future Gener. Comput. Syst.*, vol. 90, pp. 405–422, 2019, doi: 10.1016/J.FUTURE.2018.07.019.

[28] Y. Zhou, Y. Nenov, B. Cuenca Grau, and I. Horrocks, "Pay-As-You-Go OWL Query Answering Using a Triple Store," *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014, pp. 1142–1148.

[29] Y. Zhou *et al.*, "PAGOdA: Pay-As-You-Go Ontology Query Answering Using a Datalog Reasoner," *J. Artif. Intell. Res.*, vol. 54, no. 1, pp. 309–367, 2015, doi: 10.1613/jair.4757.

[30] L. Zhang, Y. Xie, L. Xidao, and X. Zhang, "Multi-source heterogeneous data fusion," *International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2018, pp. 47–51, doi: 10.1109/ICAIBD.2018.8396165.

[31] C. Hu, Y. Li, X. Cheng, Z. Liu, "A virtual dataspaces model for large-scale materials scientific data access," *Future Gener. Comput. Syst.*, vol. 54, pp. 456–468, 2016, doi: 10.1016/j.future.2015.05.004.

[32] A. Kiran and A. Rehan, "Limitations of information extraction methods and techniques for heterogeneous unstructured big data," *Int. J. Eng. Bus. Manage.*, vol. 11, pp. 1–23, 2019, doi: 10.1177/1847979019890771.

[33] T. Sakouhi and J. Akaichi, "Dynamic and multi-source semantic annotation of raw mobility data using geographic and social media data," *Pervasive Mob. Comput.*, vol. 71, 2021, doi: 10.1016/j.pmcj.2020.101310.

[34] I. Isa, L. Mark, and K. Ioannis, "A deep semantic search method for random tweets," *Online Social Networks Media*, vol. 13, pp. 1–13, 2019, doi: 10.1016/j.osnem.2019.07.002.

[35] S. Vlad and E. Martin, "Detecting Singleton Review Spammers Using Semantic Similarity," *WWW' 15 Companion*: *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 971–976, doi: 10.1145/2740908.2742570.

8

Bull. Pol. Acad. Sci. Tech. Sci., vol. 70, no. 3, p. e139434, 2022