Semantic Retrieval Method of UK Education Resource Metadata in Hierarchical Cloud P2P Network

Xia Wang and Ming-Jun Li
School of Foreign Languages, Mianyang Teachers’ College, Mianyang 621000, China
WangXia3476@163.com

Abstract. The traditional semantic retrieval method of British educational resource metadata has a low recall rate. Therefore, a semantic retrieval method of British educational resource metadata under a layered cloud peer-to-peer network is proposed. Preprocess the British education resource metadata and query and expand the semantics of the British education resource metadata. On this basis, the semantic correlation of the British education resource metadata is calculated to achieve the semantic retrieval of the British education resource metadata under the layered cloud peer-to-peer network. The experiment proves that the semantic retrieval method of the metadata of British education resources under the layered cloud peer-to-peer network designed this time has a higher recall rate than the traditional method and has practical application significance.

Keywords: Stratified cloud · Educational resources · Metadata · Semantic retrieval · Extension · Approximation

1 Introduction

Today’s society is an information-based society, the diversification of information technology to the education industry has brought new opportunities [1, 2]. As a way to obtain information, information retrieval is not only aimed at a small number of people, but also goes deep into the work, life and study of each group. With the rapid growth of educational resources on the Internet, more and more users study through the Internet. Traditional information retrieval is based on keyword matching, and the accuracy of retrieval results is low, so it will be more and more difficult for users to retrieve the required resources from massive resources [3]. How to enable users to quickly and efficiently retrieve the information that meets their needs has become an urgent problem in the field of online learning.

At present, although there are many information retrieval tools, most of them are based on keyword matching. But educational resources are interrelated, synonymous or connecting between the preceding and the following. Simply matching the words will result in a large number of resources misdetection. Traditional retrieval based on metadata semantics ignores the inherent semantic relationship between users’ retrieval
conditions and resources, resulting in information missing detection and other problems. Therefore, a semantic retrieval method of British education resource metadata based on hierarchical cloud peer-to-peer network is designed.

2 British Education Resources Metadata Preprocessing

The relationship between metadata of educational resources in Britain is analyzed before semantic retrieval of metadata of British educational resources [4]. The UK Education Resources metadata contains information as shown in the following Table 1:

| Serial number | Index                  | Content                                                                 |
|---------------|------------------------|-------------------------------------------------------------------------|
| 1             | Knowledge architecture | The overall structure including different knowledge categories and subject knowledge |
| 2             | Knowledge              | Generic term for larger areas of scientific knowledge                   |
| 3             | Subject knowledge      | Classification of knowledge according to the nature of the research object |
| 4             | Secondary subject knowledge | Can set a series of more complete curriculum knowledge according to teaching needs |
| 5             | Tertiary knowledge     | Expertise or knowledge of a degree                                       |
| 6             | Course knowledge       | Independent and complete                                                |
| 7             | Text knowledge         | Subject knowledge                                                       |
| 8             | Chapter Knowledge      | Knowledge organization                                                  |
| 9             | Unit knowledge         | Knowledge points are aggregated from a certain number of knowledge points that are highly coupled in content |
| 10            | Knowledge point        | Relatively independent and complete in the knowledge system              |

Knowledge points: in the field of educational resources, knowledge points is not divided or unsuitable meta-knowledge, in the knowledge system is relatively complete and independent of the basic unit. The relationship between knowledge points.

Implication relation: It is also called containment relation, that is, there are many small knowledge items in a knowledge. Because the knowledge points are the basic units that can not be divided or not suitable to be divided, therefore, all other knowledge levels except the knowledge points have implication relation.

Dependency: For some knowledge points, require beginners to these knowledge points have a certain sequence of learning, such as learning to multiply before must learn to add, then add and multiplication between the two points of knowledge is dependent [5]. It is precisely because of this sequential learning relationship that dependencies
are transitive. Dependencies include direct dependencies and indirect dependencies. If learning knowledge point a can direct learning knowledge point b, then both meet direct dependence relationship; If learning knowledge point a, also need to learn other knowledge points to be able to learn knowledge point b, then both meet indirect dependence relationship.

Brotherhood: As the name implies, belong to a unit of knowledge under the knowledge that is a brotherhood [6].

Parallelism: In the knowledge system, if the knowledge points with sibling relationship do not have dependencies, the knowledge points are said to have parallelism.

Reference relationship: If some knowledge points have certain relevance in the content, but these knowledge points do not necessarily have brothers or dependent relationship, if in the study of knowledge b, it may be because b and b has implication, dependence, brothers, or parallel relationship between the knowledge points can learn knowledge a, then knowledge b and knowledge a reference relationship between the existence of b. The participle structure is as follows (Fig. 1):

![Fig. 1. Segmentation module structure diagram](image)

### 3 British Education Resources Metadata Semantic Query Extension

The goal of the query extension is to find all the documents in the information collection that are relevant to the user’s query. This article focuses on expressing the information that the user needs [7]. Therefore, the concept semantic space is used in the knowledge retrieval because of its good concept hierarchy and the support of logical reasoning. In the semantic query expansion, the user will do his best to help the query. After the search, the user can obtain the various relationships between the query words through the domain concept, and then expand the concept according to the semantic relationship. Build ontology relationships as shown in the following Table 2:
Table 2. Ontology relationship building content

| Serial number | Step                                      | Content                                           |
|---------------|-------------------------------------------|--------------------------------------------------|
| 1             | Collect and organize various terms in the field | Identify the various relationships between concepts |
| 2             | Concept                                   | Analyze retrieved content                         |
| 3             | Create new project                        | Build class                                       |
| 4             | Defining attributes                       | Add various attributes and sub attributes         |
| 5             | Create instance                           | Add property values in the property pane on the right |
| 6             | Edit completed                            | Convert to other formats to save                  |

The semantic concept tree is shown as follows (Fig. 2):

![Semantic concept tree](image-url)

When a user enters a query into a search engine, it is necessary to map the query into the space of concept semantics before data retrieval, and then find out the related concepts to improve the recall rate. The details are as follows (Table 3):
Table 3. Semantic query extensions

| Index | Content                  | Method                                      |
|-------|--------------------------|---------------------------------------------|
| 1     | Formal processing        | Keyword collection                          |
| 2     | Stop word                | Remove query requests using stopwords       |
| 3     | Query expansion          | Semantic data retrieval                     |
| 4     | Extract keywords         | Formal processing                           |
| 5     | Concept map              | Node matching                               |
| 6     | Building search tree judgments | Exact matching of each node                  |
| 7     | Building the concept tree | Close to the initial intent of the user’s query |
| 8     | Concept expansion        | Setting thresholds for direct expansion     |
| 9     | Keyword merge            | Add concepts greater than a threshold to the query expansion term set |

4 Semantic Relevance Calculation of UK Education Resources Metadata

Semantic similarity is limited by people’s knowledge structure. It is a kind of concept with strong subjective concept. It is difficult to define semantic similarity without actual situation [8]. The relationship between words and phrases is very complicated. Usually, a specific number is used to express the similarity between words and phrases. If the possibility of not changing the text between words and phrases is less, then the similarity between two words and phrases is said to be less. Its resource structure table is shown in the following Table 4:

Table 4. Resource structure content

| Educational Resources | Content                | Data element       |
|-----------------------|------------------------|--------------------|
| Resource 1            | Media material         | Image material     |
| Resource 2            | Test questions         | Audio material     |
| Resource 3            | Test paper             | Video footage      |
| Resource 4            | Courseware             | Animation material |
| Resource 5            | Reference              | Case               |
| Resource 6            | Online Course          | Index              |
| Resources 7           | Frequently Asked Questions | Question Bank   |
| Resources 8           | Text material          | Answer             |

If a word cannot be exchanged 100% between words in a context, the similarity of the two words is 0, and the similarity of a word to itself is between 0 and 1. The shortest
path between two concepts in the semantic tree is the semantic distance. The smaller the semantic distance between concepts, the greater the similarity between them. The influence of semantic distance on semantic similarity can be expressed as follows:

\[ S = \frac{az}{b} \]  \hfill (1)

In the formula, \( b \) represents the distance between the two conceptual nodes and all the corresponding distances on the edge of the connected path, such as the distance between parent and subclass 1; \( a \) represents a similarity calculation parameter; \( z \) represents the semantic distance between concepts and concepts.

Semantic coincidence refers to the same proportion of the upper concept in the path between the two concept nodes in the tree graph, that is, the number of common nodes of the upper node/the number of summary points. The higher the semantic overlap between the two concepts, the higher the degree of similarity between them. The effect of semantic coincidence on semantic similarity is mainly expressed as:

\[ D = \frac{s}{Wh} \]  \hfill (2)

In the formula, \( s \) represents the collection of the number of nodes passing through the node to the root; \( W \) represents the intersection of the number of nodes passing through the root node; \( h \) represents the union between the number of nodes passing through the node to the root.

Node density refers to the number of child nodes owned by a conceptual node. If the number of children a concept node has is greater, then the node density is greater. The density of the tree here refers to the number of branches of the nodes. If the node density of the common ancestor node between the two conceptual nodes is greater, then the similarity between the two nodes is greater. The effect of node density on semantic similarity can be expressed as:

\[ j = \frac{d}{d(T)} \]  \hfill (3)

In the formula, \( j \) represents the closest common coincidence between concept nodes and concept nodes; \( d(T) \) represents the length of the concept tree; \( d \) represents the degree of proximity of the two nodes at the level.

Hierarchy. In domain ontology, for two nodes that are not synonymous, the order of the hierarchy of nodes will also have an impact on the similarity of the nodes. The order of the hierarchy has an impact on the conceptual similarity:

\[ F_i = (S_i + q)H \]  \hfill (4)

In the formula, \( S_i \) indicates the level at which the node is located; \( q \) represents the lowest level in the domain ontology; \( H \) represents the similarity between semantics.

In the final phase of semantic similarity calculation, the calculation of similarity between concepts is affected by the weight of query words. In this paper, the threshold value should be set dynamically. The concept of high query term weight should be closer to the user’s intention. In order to search more resources, the similarity threshold value
should be lower, and the concept of low query term weight should be higher [9]. After
illustrating the various factors that affect similarity, we use examples to illustrate the
calculation process of semantic relevance. Among many knowledge points and the tree
graph constructed by knowledge points, the relationship is relatively complex. When
calculating semantic similarity, the relationship between nodes is expressed.

5 Implementation of Semantic Retrieval for Metadata of British
Education Resources

On the basis of the above-mentioned UK education resources metadata preprocessing,
UK education resources metadata semantic query expansion and UK education resources
metadata semantic relevance calculation, the UK education resources metadata semantic
retrieval process is shown in the following Fig. 3:

![Diagram of semantic retrieval process]

**Step 1**: Client $X$ can initiate resource discovery request $Q$ to any resource node $N$;
**Step 2**: Parse the query, and the learning object metadata standard basic pattern
fragments are shown in the following Table 5:

| Extract core concept set $c_i$, core concept constraint $P(c_i)$, and extension layer concept set $c_i$; |
| --- |

**Step 3**: Based on the community resource localization, forward the query $Q$ among
the different communities $Q$ to determine whether the $c_i$ corresponds to the community
constraint $P(c_i)$ is compatible with $P(K)$, that is, whether the instance satisfying the
two constraints is intersected, the existence traverses all nodes in the community, returns
the resource object $id$ satisfying $P(K)$ to $N_i$, and $N_i$ returns the result to the client $X_i$. 
Table 5. Learn object metadata standard basic pattern fragments

| Numbering | Name            | Explanation                                                                                   |
|-----------|-----------------|---------------------------------------------------------------------------------------------|
| 1         | Universal       | This category describes some general information about learning objects                     |
| 1.1       | Title           | Learning Object Name                                                                        |
| 1.2       | Key words       | Keywords or phrases describing the subject of the learning object                          |
| 1.5       | Education       | Some key characteristics of learning objects in education and teaching                      |
| 5         | Resources       | Specific types of learning objects                                                          |
| 5.2       | Resource Type   | Classification path of learning objects in a specific classification system                 |
| 9.2       | Path            | …                                                                           |

or passes the query to the next community. Until a specific TTL value is reached or all communities are traversed;

Step 4: Do a local match for the local index of the term resource and rate the match. The UK educational resource organization is shown below (Fig. 4):

Fig. 4. Organisational structure of UK educational resources
The query $Q$ is sent to the resource Agent node responsible for these resource objects to determine whether it is compatible with the corresponding node community constraints, otherwise the operation will end if it is not compatible, and the matching algorithm is:

$$D = \left( \sum E + \frac{1}{2} \sum I \right)^2$$  \hspace{1cm} (5)

In the formula, $I$ is the British Educational Resources Metadata; $E$ represents semantic relevance.

Finally, the resource object with matching degree meeting a specific threshold is returned to $N_i$ and finally to client $X$.

To avoid repeated processing of the same query and to alleviate the hot issues of some popular queries, the query results can cache a specified time on the query initiation resource agent node and the query delivery routing node.

At the same time, the load balance should be ensured in the semantic retrieval of the UK educational resources. Even after the corresponding processing of some UK educational resources metadata semantic retrieval, the conceptual terms of the index in each engine node cannot be completely balanced, and the number of resources of some engine node indexes may be large [10]. The core metadata is therefore addressed in a community-based overlapping network approach to meet specific constraints, as follows (Table 6):

| Serial number | Content                     | Method                                    |
|---------------|-----------------------------|-------------------------------------------|
| 1             | Concept element             | To divide the node space                  |
| 2             | Community logo              | Community entrance constraints            |
| 3             | A new resource object       | Added to affiliated resources             |
| 4             | Metadata parsing            | Extract the core concept and extended     |
|               |                             | concept elements for processing separately|
| 5             | Core-level metadata         | Connectivity between different communities|
| 6             | Passed to all communities in the system | Match the community |

At the same time, the traffic of different terms is unbalanced, so the computing load of some engine nodes may be too high, which affects the overall performance of metadata semantic retrieval of educational resources in Britain. To solve this problem, this paper adopts an adaptive load balancing mechanism, which is expressed as:

$$D = \frac{m}{qx}$$  \hspace{1cm} (6)

In the formula, $D$ represents the number of resources for the engine node index; $m$ represents the index key value range; and $x$ represents the index data quantity or per unit time data access.
According to the above process, the metadata semantic retrieval of British educational resources under hierarchical cloud peer-to-peer network is completed.

6 Experiments

In order to verify the validity of the UK education resources metadata semantic retrieval method based on layered cloud peer-to-peer network, experiments are carried out to compare the recall of the two methods.

6.1 Experimental Evaluation Indicators

The evaluation of retrieval performance is mainly focused on whether the retrieval results are satisfactory after the user inputs the retrieval information. Recall is not only a key index in the evaluation, but also a key index in the performance evaluation of traditional information retrieval.

Recall is the ratio of the relevant literature to all the relevant literature in the search results after the user enters the search term, that is, the success of finding the literature related to the search from all the literature. Recall is the degree of the relevant results retrieved. The formula for recall rate shall be:

\[ R = \frac{F}{S} \times 100\% \]  

In the formula, \( S \) represents the amount of all relevant resources; \( F \) represents the number of resources retrieved.

The diagram shows (Fig. 5):

\[ A \cap B \]

Fig. 5. Schematic illustration of recall
6.2 Experimental Content

What this experiment uses is “the computer network” in the curriculum each discrete knowledge spot, including has the text, the audio frequency, the video frequency, the webpage and so on each kind of material. The experiment selects more than 100 relatively huge knowledge points, more than 1000 “computer network” course articles. Among them, knowledge points include WAP, SNMP, asynchronous transmission mode, conflict detection, VPN and so on. This paper has carried on several groups of query tests to these complex knowledge points.

In this test, this paper uses the artificial evaluation method to evaluate the relevant literature in the experimental results, and the final retrieval results are judged by artificial correlation.

6.3 Experimental Platform

The experimental platform used in this experiment is shown in the following Fig. 6:

![Experimental platform diagram]

Fig. 6. Experimental platform

6.4 Experimental Result

On the basis of the above experimental preparation, the structure of the experimental results is shown in the following Fig. 7:
Analysis of the above Fig, in the semantic retrieval of British education resource metadata under the peer-to-peer network, the overall query recall rate of traditional methods is lower than that of this design method. It can be seen that this design method has a higher recall rate. This is because the method designed this time analyzes the semantic similarity in detail, thus ensuring a high recall rate.

Therefore, through the above experiments can prove that the design method is effective, with practical significance.

7 Conclusion

In order to solve the problem of low recall rate of the semantic retrieval method of British education resource metadata, a semantic retrieval method of British education resource metadata under a layered cloud peer-to-peer network is proposed. By preprocessing British educational resource metadata, querying and expanding the semantics of British educational resource metadata, and calculating the semantic relevance of British educational resource metadata, the semantic retrieval of British educational resource metadata under a hierarchical cloud peer-to-peer network can be achieved. The experimental results show that the designed semantic retrieval method of British educational resource metadata under the hierarchical cloud peer-to-peer network has a high recall rate and has practical application significance.

Fund Projects. Special project to build mianyang normal university in 2019: Theoretical research and practical exploration of foreign language teacher education (project number: 2019 mysytz11) college.

References

1. Fan, Y.X., Guo, J.F., Lan, Y.Y., et al.: A context-aware deep sentence matching model. J. Chinese Inf. Process. 31(5), 156–162 (2017)
2. Chen, Q., Dai, Q.: Research and Application of Semantic Understanding Methods in Searching Advertising

3. Microcontrollers Embedded Syst. 19(6), 13–17 (2019)

4. Miao, T.P., Han, J.J., Wang, Z.J.: Intelligent semantic understanding of the morphological characteristics of Chinese function words in search engines. Outlook Electron. Technol. 26(6), 52–55 (2019)

5. Zhang, S.W., Ouyang, C., Yang, X.H., et al.: Word semantic similarity computation based on integrating HowNet and search engines. J. Comput. Appl. 37(4), 1056–1060 (2017)

6. Sun, M.Y., Tian, X.D.: Similarity sorting method for retrieval results of linear algebra formula. Comput. Eng. 44(4), 253–261 (2018)

7. Zhang, Q.Y., Lin, M., Zhang, S.J.: Research on automatic construction of domain concepts on Wikipedia semantic knowledge base. Appl. Res. Comput. 35(1), 130–134 (2018)

8. Han, Y.Q.: The realization and mechanism of library subject collection semantization based on the entity-relationship mapping method of BIBFRAME model: take “information retrieval” as an example. Libr. J. 36(9), 35–41 (2017)

9. Feng, S.X., Zhang, Y.M.: Study on semantic retrieval to improve IT level of Northeastern Asian shipping center. Logistics Technol. 36(7), 150–153 (2017)

10. Wu, X., Zhou, D.: Personalized query expansion method based on multiple semantic relationships. Pattern Recogn. Artif. Intell. 30(11), 1039–1047 (2017)

11. He, Y., Li, T., Wang, W., et al.: A semantic similarity integration method for software feature location problem. J. Comput. Res. Dev. 56(2), 394–409 (2019)