Research on Reasoning and Retrieval Methods Based on Mongolian Curriculum Areas of Semantic Web

Hongbo Chen  
College of Information Engineering Inner Mongolia University of Technology Hohhot 010051, P.R.China  
Email: chb263@163.com  

Yila Su and Chuan Fan  
College of Information Engineering Inner Mongolia University of Technology Hohhot 010051, P.R.China  
Email: suyila@tsinghua.org.cn

Abstract—The backwardness of the Mongolian network teaching resources results in its low reuse rates and utilization. For this situation, a retrieval method of semantic web based on Mongolian curriculum areas was set up. Firstly, the method established the Mongolian ontology of course ‘Artificial Intelligence’ in area of teaching, it uses a relationship database MySQL to record ontology information, on the top of that, it uses semantic similarity calculation and reasoning rules established to do the semantic information reasoning, in the end, it builds a query and retrieval prototype system based on the field of Mongolian teaching. Experimental results show that this method can effectively improve the information retrieval recall and precision rates, and it established Mongolian teaching resources which can be shared and interoperable knowledge hierarchy.

Index Terms—Mongolian, semantic Web, ontology, reasoning, curriculum, information retrieval

I. INTRODUCTION

As traditional information retrieval technology is mainly based on keyword matching, syntactical level search, it was lack of processing and presentation in the semantic level, the search results are incomplete and inaccuracy issues. However, solve this problem is from the simple keyword matching converted to semantic matching in the search process. Semantic Web[1][2] is proposed for traditional information retrieval provides a new technique that enables the machine to be understood semantic information. In the information age with rapid development of Internet, online teaching more and more welcomed by the people, which had been accepted and become a new trend at domestic and abroad. This new way of teaching not only make the knowledge and information spread fast, but also provide a platform for the sharing of teaching resources[3]. But in the network exist the large number of teaching and learning resources, mostly using simple semantic knowledge of the hierarchy without sharing and interoperability, so exists lower reuse and utilization rates issues. Applying the ontology technology to the field of education started in the early twenty-first century, it was university of Kalsruhe in Germany that constructed an ontology building tools KAON combined semantic web with E-learning[4][5], Stanford University analyzes the differences between the traditional network of teaching and teaching based on semantic WEB networks, and it proposed ontology-based metadata that describe the learning resources, it builds a ontology-based courseware resource editor [6]. In China, the ontology research started in early 2002, it was Hui Dong in university of Wuhan, ontology application in digital libraries research [7], and Junfeng Song, Weiming Zhang, ontology-based information retrieval model [8], who is the scholar of engaging in applying the ontology theory to the field of teaching area. Mongolian language [9] [10] is one of the official language in Inner Mongolia, Mongolian computer information processing system was set up by the1980s, which provide good conditions for the promotion and application of computer processing of Inner Mongolian information. Inner Mongolia University, Inner Mongolia Academy of Social Sciences, Inner Mongolia Normal University and other units have developed a Mongolian computer information processing systems. Mongolian international standard code achieved recognition of the National Organization for Standardization and UNICODE Technical Committee. Because of poor conditions, small-scale schools, distribution of concentration issues in Mongolia, led to low levels of Mongolian teaching, greatly influenced the Mongolian teaching progress. Under these conditions, developing Mongolian network teaching can greatly improve.

II. ESTABLISHMENT OF "ARTIFICIAL INTELLIGENCE" CURRICULUM MONGOLIAN ONTOLOGY

A. The Needs Analysis of Mongolian Curriculum Areas Ontology

The needs analysis[11] is the cornerstone of building knowledge domain ontology[12][13], after understanding the specific objectives and characteristics in knowledge domain ontology, it can be normalized, detailed and
specific described for ontology. So, in the beginning of building Mongolian curriculum domain ontology, the needs analysis is essential. According the specific purpose and significance of Mongolian curriculum areas ontology mentioned in this paper, specified information is as followed:

First, the purposes of building Mongolian curriculum ontology, ontology used in Mongolian curriculum areas construction, is mainly to achieve a common understanding in a hierarchy of knowledge of Mongolian specific curriculum areas, and meet the knowledge sharing, knowledge representation and knowledge inquiry between different systems, by adding semantic information in curriculum ontology can greatly improve the retrieval recall and precision rates.

Second, the Mongolian curriculum domain ontology covered subject areas. In order to avoid gaps in knowledge and meet information interaction between the various cross-disciplinary, in building ontology should take into account all the knowledge points of the curriculum and cross-disciplinary knowledge.

Third, the basic users of Mongolian curriculum areas ontology are teachers, students, and Mongolian researchers.

Fourth, the schedule of constructing the Mongolian curriculum areas ontology is that taking the particularities and knowledge Mongolian comprehensiveness, complexity, and other factors and the rules of practical application and effectiveness into account, we need to allocate time appropriately according to the knowledge level of difficulty.

Fifth, choose a correct description language when building Mongolian curriculum ontology. In order to have a good expression and expansion capabilities in the field of knowledge, we use the OWL ontology language.

B. Establish to Core Ontology Library of Mongolian Teaching Field

First, we should get concepts and relationships between concepts in the field of knowledge in the beginning of building Mongolian curriculum areas ontology, and define the corresponding class. We can through the network, books, authoritative experts and existing ontology to get the organizational structure knowledge areas, according to the organizational structure of the field of knowledge and various properties of concepts to get the top of concepts, one-level concepts and so on. Each of concepts inherits the basic attributes of parent class, and gradually form a hierarchical model. Concept in each layer can be abstracted as the ontology’s basic classes, which subclass should inherit all the properties of the parent class, and subclass should have a new property. In this paper, combined with knowledge structure and teaching objectives of "artificial intelligence" curriculum, defines four top classes, include artificial intelligence research focus, applications, search reasoning, knowledge representation. The first layer of subclass more than 20, The second layer of subclass more than 100, coverage all the knowledge points of "Artificial Intelligence (fontsize 10pt "Artificial Intelligence")".

Secondly, according to different classes to determine the class attributes. In this paper, according to different relationships between curriculum resources and combination of teaching objectives and tasks, which defined attributes as following table:

**TABLE I.**

**THE MAIN ATTRIBUTES OF CURRICULUM AREAS ONTOLOGY**

<table>
<thead>
<tr>
<th>Data attribute name</th>
<th>Property description</th>
</tr>
</thead>
<tbody>
<tr>
<td>共识</td>
<td>Author</td>
</tr>
<tr>
<td>语义</td>
<td>Language</td>
</tr>
<tr>
<td>适配日期</td>
<td>Create date</td>
</tr>
<tr>
<td>使用角色</td>
<td>Using Roles</td>
</tr>
<tr>
<td>难度</td>
<td>Difficulty</td>
</tr>
<tr>
<td>关键词</td>
<td>Keyword</td>
</tr>
<tr>
<td>知识同义词</td>
<td>Knowledge synonyms</td>
</tr>
<tr>
<td>关键元素 (Y/N)</td>
<td>Key elements (Y/N)</td>
</tr>
<tr>
<td>教学目标</td>
<td>Teaching objectives</td>
</tr>
<tr>
<td>知识点重要度</td>
<td>The size of knowledge</td>
</tr>
</tbody>
</table>

**TABLE II.**

**THE OBJECT PROPERTIES OF CURRICULUM AREAS ONTOLOGY**

<table>
<thead>
<tr>
<th>Object attribute name</th>
<th>Attribute description</th>
<th>Attribute explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>父类知识</td>
<td>Parent knowledge</td>
<td>A knowledge point corresponds to a parent knowledge</td>
</tr>
<tr>
<td>子类知识</td>
<td>Children knowledge</td>
<td>A knowledge point has one or more sub-knowledge</td>
</tr>
<tr>
<td>引用</td>
<td>Refer</td>
<td>Knowledge points refer to other resources</td>
</tr>
<tr>
<td>引用者</td>
<td>Referred</td>
<td>Knowledge points is refered by other resources</td>
</tr>
<tr>
<td>引用文章</td>
<td>Quote</td>
<td>Articles refer to other resources</td>
</tr>
<tr>
<td>引用文章</td>
<td>Quoted</td>
<td>Articles is refered by other resources</td>
</tr>
<tr>
<td>共享引用</td>
<td>Common reference</td>
<td>Two or more articles refer to the same resource</td>
</tr>
<tr>
<td>共享引用</td>
<td>Common reference</td>
<td>Two or more articles are refered by the same resource</td>
</tr>
</tbody>
</table>

Third, combined with teaching objectives of "Artificial Intelligence" curriculum describe the logical relationships between different concepts. This paper defines six basic relationships described curriculum ontology, including...
the relationship between knowledge and chapters, the relationship between knowledge and resources, the upper relationship between knowledge, the lower relationship between knowledge, the parity relationship between knowledge, the relationship between knowledge, which defined relationship as following table:

### TABLE III

**THE BASIC LOGICAL RELATIONSHIPS BETWEEN CLASSES**

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Relationship description</th>
</tr>
</thead>
<tbody>
<tr>
<td>上下关系</td>
<td>relationship between knowledge and chapters</td>
</tr>
<tr>
<td>左右关系</td>
<td>relationship between knowledge and resources</td>
</tr>
<tr>
<td>前后关系</td>
<td>upper relationship between knowledge</td>
</tr>
<tr>
<td>左右关系</td>
<td>lower relationship between knowledge</td>
</tr>
<tr>
<td>左右关系</td>
<td>parity relationship between knowledge</td>
</tr>
</tbody>
</table>

Fourth, build instance for each specific curriculum class.

### III. SEMANTIC SIMILARITY CALCULATION

Semantic similarity\[14\][15][16] definition is two or more different concepts which have the same or similar feature, we then define these concepts are similar, for example, using Sim (C1, C2) indicates similarity between C1 and C2, the similarity calculation should satisfy the following situations:

1. The similarity of Sim (C1, C2) has a value is any real number in \([0, 1]\).
2. When the concept of C1 and C2 have same attribute characteristics, namely C1 = C2, then Sim (C1, C2) = 1.
3. When the concept of C1 and C2 have completely different attribute characteristics, then Sim (C1, C2) = 0.
4. The relationship is symmetrical between concept of C1 and C2, namely Sim (C1, C2) = Sim (C2, C1).

The impact factors of semantic similarity calculation\[17\][18] include semantic distance, concept node depth, concept node density and adjusting parameters. Depending on different impact factors have different influence level, summed similarity is calculated as follows:

\[
\text{Sim}(C1,C2) = (\alpha / \text{Dis}(C1,C2)) + (\alpha + \beta + \gamma) \left( \frac{\text{Depth}(C1) + \text{Depth}(C2)}{|(\text{Depth}(C1) - \text{Depth}(C2)| + 1} \right)^\delta, \text{Density}(C1, C2)
\]

\[
\text{Dis}(C1,C2) = |\text{Depth}(C1) - \text{Depth}(C2)| + 1
\]

In the above formula, \(\alpha / \text{Dis}(C1,C2)\) + represents semantic distance can influence the semantic similarity calculation, since semantic distance largely affected the semantic similarity calculation, so \(\alpha\) has the larger weights; \((\text{Depth}(C1) + \text{Depth}(C2)) / (|\text{Depth}(C1) - \text{Depth}(C2)| + 1)\) represents concept node depth less affected the semantic similarity calculation, since concept node depth less affected the semantic similarity calculation, so \(\beta\) has the smaller weights; \(1 / \text{Density}(C1, C2)\) represents concept node density can influence the semantic similarity calculation, since concept node density less affected the semantic similarity calculation, so \(\gamma\) has the smaller weights, and \(\alpha + \beta + \gamma = 1\).

In this paper, extract part of a collection ontology to calculate semantic similarity. Extract part of a collection of ontology \([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]\) denote \{\text{(Search and Reasoning)}, \text{(And or Graph)}, \text{BP}, \text{(Production system)}, \text{(Search and Reasoning)}, \text{(Heuristic search)}, \text{(Blind search)}, \text{(State graph search)}, \text{(Reverse reasoning)}, \text{(Bidirectional reasoning)}, \text{(Etc Search)}, \text{(Exhaustive Search)}, \text{(Breadth-first search)}\}, as shown below:

![Figure 1. Part of a collection of ontology](image)

This paper take \(\alpha = 1, \beta = 0.9, \gamma = 0.05\), the similarity between each node has been calculated as follows:

### TABLE IV

**SEMANTIC SIMILARITY VALUE OF EACH NODE**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.576</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
</tr>
<tr>
<td>2</td>
<td>0.576</td>
<td>1.0</td>
<td>0.391</td>
<td>0.550</td>
<td>0.550</td>
<td>0.550</td>
<td>0.550</td>
</tr>
<tr>
<td>3</td>
<td>0.391</td>
<td>0.391</td>
<td>1.0</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
</tr>
<tr>
<td>4</td>
<td>0.368</td>
<td>0.368</td>
<td>0.399</td>
<td>1.0</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
</tr>
<tr>
<td>5</td>
<td>0.550</td>
<td>0.550</td>
<td>0.399</td>
<td>0.293</td>
<td>1.0</td>
<td>0.293</td>
<td>0.293</td>
</tr>
<tr>
<td>6</td>
<td>0.398</td>
<td>0.398</td>
<td>0.398</td>
<td>0.293</td>
<td>0.293</td>
<td>1.0</td>
<td>0.293</td>
</tr>
<tr>
<td>7</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>0.293</td>
<td>1.0</td>
</tr>
<tr>
<td>8</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
<td>0.368</td>
</tr>
<tr>
<td>9</td>
<td>0.391</td>
<td>0.391</td>
<td>0.391</td>
<td>0.391</td>
<td>0.391</td>
<td>0.391</td>
<td>0.391</td>
</tr>
<tr>
<td>10</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
</tr>
<tr>
<td>11</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
</tr>
<tr>
<td>12</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
</tr>
<tr>
<td>13</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
</tr>
<tr>
<td>14</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
<td>0.399</td>
</tr>
</tbody>
</table>

© 2014 ACADEMY PUBLISHER
User query and results display interface: This function is information exchange between user and the system, namely provide users with entered interface and returns the query results.

Ontology information database: Ontology as the core of semantic retrieval, it defines the concept of knowledge, relationships, and a collection of entities and attributes, so that people can reach for a common understanding of knowledge areas, and achieve information sharing and reuse, so building and managing ontology is very crucial. Ontology information mainly from books, experts and have defined the concept to get, it major complete semantic expansion and semantic annotation.

Web information management: This paper mainly achieved query and retrieval the resources on the Web, in order to ensure precision rates we must have sufficient resources of Web information are available for retrieval. In this function module, ontology repository add Web information which has been collected, can also add some Word, PDF documents, which will help the integration of information resources.

Semantic extension and reasoning[19][20]: This function module implements when user enter keyword retrieving information, according to ontology library information which has been built, we should map original query keyword and ontology library information to conduct semantic similarity calculate and semantic extensions, at the same time according to inference rules conduct semantic information inference, which get the ontology information similar to original query words, added new query words to conduct retrieval, so improve the retrieval accuracy. In this paper, semantic extension mainly include synonyms extensions, contain extensions, hypernym extensions, parity word extensions, hyponym extensions and other extensions. Ontology reasoning according to inference rules apply relevant inference engine analysis ontology OWL document, this paper mainly involves some relevant curriculum knowledge reasoning.

C. Add Inference Rules

Jena has ontology query and reasoning mechanisms, if achieving Jena reasoning, we have to formulate inference rules, the inference engine according to inference rules conduct reasoning, and get a collection of concept which associated with given the concept. In this paper, in order to achieve the "Artificial Intelligence" curriculum knowledge areas ontology query and retrieval, practical application and effectiveness must be given as followed:

(1) As in object-oriented learning process has given class inheritance, so the class inheritance concept is also applicable to ontology, the class inheritance inference rules are as follows:


Note: if c1 is a subclass of c2 and p is an entity or attribute of c1, then p also belongs to an entity or attribute of c2.

(2) In the ontology concept, if two classes are the same class, the two classes have the same attribute value, inference rules are as follows:

Note: in the Rule 2 and Rule 3, c1 and c2 are the same class, then c1 and c2 have the same attribute values.


(3) Classes’ contain relationships, inference rules are as follows:


Note: if c1 is a subclass of p and p is a subclass of c2, then c1 is c2 subclass.

Rule 5: (? C1, owl: SameAs, ? C2), (? C1, p# ConnectWith, ? C3) -> (? C2, p# ConnectWith, ? C3)

Note: c1 and c2 are synonymous class, c1 and current class have the same characteristics p, then c2 and c3 also have the same characteristics p.

(4) Citing and cited, the inference rules are as follows:

Rule 6: (? C1, # citing, ? P), (? C2, # citing, ? P), notEqual (? C1, ? C2) -> (? C1, # bothciting, ? C2)

Note: c1 class c2 cite resource p, c1 and c2 are different, then c1 and c2 simultaneously cite p.


Note: p cite resources c1 and c2, c1 and c2 are different, then c1 and c2 are simultaneously cited by p.

D. Storage Curriculum Ontology Information

Jena provide some interfaces which RDF data can storage relational database, the interfaces used to access and maintain RDF data, and the interface includes Model, Resource and Query and so on. Jena also provide some program interfaces which can support relational database, including MySQL, HSQLDB, SQL Server, etc. This paper will use a relational database MySQL storage ontology data, the main steps include:

1. Load the database JDBC driver;
2. Create a database connection;
3. Create a ModelMaker database model for the database;
4. Create a model for ontology information.

E. Design and Achievement of Semantic Retrieval Based on Mongolian Curriculum Areas

1. This design based on a curriculum areas of Mongolian semantic retrieval system is based on B / S structure, achieve Mongolian semantic retrieval priority is display properly Mongolian characters on the browser, due to most browsers do not support the Mongolian characters, therefore, in this paper, the design is required before download and install IE Tab Multi plug to solve Mongolian not display correctly on the browser issue.

2. The curriculum areas Mongolian semantic retrieval main purpose is that when we input knowledge point need to retrieve, click submit button, the system will conduct semantic similarity calculation and semantic expansion according to keyword, we will get the ontology information similar to original query words, added new query words to conduct retrieval, so improve the retrieval accuracy.

3. In search module, it has not only the basic information and meanings of search words, more importantly, it also has the all the semantic information corresponding to the search words by the system of semantic query expansion. For instance, given the point of search word "үйлэх өрөө (and or graph)", we achieve semantic relationships through semantic query expansion: the relationship of knowledge and chapters, the relationship between knowledge and resources, knowledge, relationships between the upper, lower, relationship between knowledge, knowledge of the relationship between the parity, the relationship between knowledge points. We use this information to call the corresponding class in order to find the specific semantic relations under the relevant knowledge, including the author, creation date, language, size, using roles, synonyms, parity words, the upper word, knowledge point range, whether it is key elements, relevant content, degree of difficulty, teaching objectives, citation, cited the same arguments as well as with other. Specific semantic query flow is as follows:

V. CONCLUSIONS

This paper according to Mongolian network teaching resources exist the information can not be shared and non-interoperable issue, we study and design information retrieval system based on Mongolian curriculum areas semantic WEB. The main research work is as follows:

1. This paper carried out a detailed analysis for "artificial intelligence" curriculum, we established curriculum ontology based on Mongolian and
This paper analyzes the impact factors of semantic similarity calculation, extracted part of a collection of ontology to calculate semantic similarity, we get semantic similarity of each knowledge points.

By using the MySQL relational database storage "Artificial Intelligence" curriculum ontology information.

Designed and Achieved retrieval system based on Mongolian curriculum areas Semantic Web.

ACKNOWLEDGMENT

This work is partially supported by National Natural Science Foundation of China (61363052), Natural Science Foundation of Inner Mongolia Autonomous Region (2012MS0904), College and University Scientific Research Project of Inner Mongolia Autonomous Region (NJZZ12046) and Inner Mongolia Science and Technology Plan Project (20130362), Foundation of Inner Mongolia University of Technology (ZD201001).

REFERENCES


