Ontology and CBR-based Dynamic Enterprise Knowledge Repository Construction

Huiying Gao
School of Management and Economics, Beijing Institute of Technology, Beijing, China, 100081
Email: huiying@bit.edu.cn

Xiuxiu Chen
School of Management and Economics, Beijing Institute of Technology, Beijing, China, 100081
Email:xiuxiuchen77@163.com

Abstract- **The efficiency of knowledge sharing and learning is the key to obtain sustainable development for the knowledge-intensive industry. However, current application of enterprise knowledge repository can hardly adapt to the personalized retrieval with semantic expansion and can not support the dynamic mechanism of knowledge sharing. This paper focuses on an integrated framework and operating processes of dynamic knowledge repository construction. Through analyzing the key technology points of business logic processing layer and data services layer particularly, the ontology and CBR-based knowledge storage and retrieval mechanism are studied, which improve the effectiveness of knowledge management.**

Index Terms—**ontology; case based reasoning; knowledge repository; semantic retrieval**

I. Introduction

With the development of knowledge economy and the superheating competition of the market, many knowledge-intensive industries such as aviation, advanced manufacturing, IT and consulting are suffering from the distress of increasing knowledge assets outflow, where technical requirements span is wide and management task is difficult and complicated. The recent research paper in the Journal of Knowledge Management shows that knowledge application is directly related to organizational performance [1]. A good environment of knowledge sharing and learning will improve the enterprise′s efficiency and service; finally make it get sustainable competition.

However, the most precious knowledge of the enterprise often exists in the minds of its employees, work processes, experiences, and in electronic or written form, etc. The research paper published in the Information Systems Research in 2010 pointed that knowledge capabilities with IT contribute to firm innovation [2], and also demonstrated that individuals will absorb and recreate knowledge in the organization with strong knowledge storage and retrieval capabilities [3].Therefore, the knowledge repository construction will play a vital role in the knowledge transformation and sharing between personal and the enterprise.

The massive domain knowledge resources and experiences are accumulated in many knowledge-intensive enterprises in recent years. However, there are still many urgent problems that they have to face. There is lack of general knowledge model for domain support, and there is lack of semantic support for the knowledge case representation, retrieval and reuse. The organization of knowledge repository fixes in a single form and the hierarchical structure is ambiguous. Besides, the knowledge repository with weak case learning ability cannot meet the dynamic mechanism of knowledge reuse.

This paper aims at a dynamic enterprise knowledge repository construction on ontology and case-based reasoning (CBR). At first in section II the ontology and CBR are briefly introduced and the state of art is summarized. After that in section III the framework of ontology and CBR-based dynamic enterprise knowledge repository is proposed and the key technology points of business logic processing layer and data services layer are described respectively. Take an information system consulting company as an example, in section IV, the process of the enterprise domain ontology is shown and the dynamic mechanism of case repository is designed in section V, including the case representation, organization, retrieval and learning. Finally section VI displays our conclusion and outlook.

II. Onrology And Case-Based Reasoning

A. Ontology

Ontology is the formal clear specification of the sharing conceptual model; it captures the basic domain terms and their relationships, defines the relevant rules to determine the vocabulary extension, and finally forms a knowledge structure model in specific areas in order to achieve the consistent understanding of the domain knowledge [4]. As ontology provides a clear semantic

This work was supported in part by the National Natural Science Foundation of China under Grant 71102111 and Beijing Institute of Technology under Grant 3210050320908.
Corresponding author：Huiying Gao Email: huiying@bit.edu.cn

doi:10.4304/jsw.7.6.1211-1218

and knowledge description of the concepts and interrelation, it can be adapted to the cases despription and hierarchical structure storage and support the semantic knowledge retrieval.

B. Case-based Reasoning

CBR is an important problem solving and learning method based on knowledge in the artificial intelligence field. It has good extensibility and the ability to learn [5]. Each processed problem is described as the feature set and solutions, then stored as a case in the system.When the new problem comes, the most similar case will be retrieved and modified if necessary. The modified cases will be seen as an new case and stored in order to realize the reuse and relearning of the cases. Also , case retrieval will be the key point of the case reasoning.

C. Related Research

Dynamic knowledge repository management is a hot issue in the current information field. Many scholars have done a great deal of researches based on ontology.

In the Journal of Information Science, many related researches are delivered such as knowledge extraction [6], the uniform knowledge representation [7], the knowledge matching and retrieval [8] with ontology technology and so on. Moreover, in the Journal of Knowledge Management, information processing based on ontology in the construction process of the knowledge management system is explored [9]. Besides, Liao Liangcai imported ontology into the knowledge management system and realized the enterprise knowledge management through the semantic expansion, reasoning and retrieval finally [10]. All these researches show the important role of ontology in the realization of knowledge sharing and reuse.

Meanwhile, the unstructured knowledge such as experiences and minds of employees is more suitable to store in the form of case, and it is easier to realize the dynamic knowledge management based on CBR. So the following research focuses on dynamic enterprise knowledge repository model with the effective combination of ontology and CBR technology. With the establishment of domain ontology, the semantic consistency for knowledge representation and storage is ensured and semantic expansion of the user′s query demand is realized. Besides, case modification, learning, reuse and new case formation enhance the adaptability of knowledge repository and realize its dynamic construction.

III. Ontology And CBR-Based Dynamic Enterprise Knowledge Repository Framework

A. Integrated Framework

An ontology and CBR-based dynamic enterprise knowledge repository model is discussed in this part. The basic framework is illustrated in Fig.1, which adopts a four-tier system structure: customer application layer, business logic processing layer, data services layer and system layer. All these different layers work closely together to complete the work of knowledge management. The main functions of each layer are described as follows.

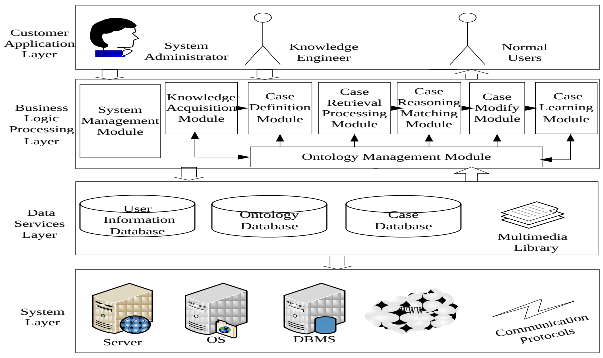

Figure1. The framework of dynamic knowledge repository

The customer application layer provides a good interaction interface for the users including knowledge engineers, normal users and the system administrator. The business logic processing layer encapsulates the core functional modules of the knowledge repository system, which are responsible for knowledge acquisition, representation, case definition and storage, ontology analysis, as well as case retrieval, learning and so on. The data service layer is the basic part of this system, which logically realizes the expression and organization of the user information database, multimedia database, ontology database and case database. The system layer aims at offering operation system, database management system, Server, data standards, network, communication protocols, and many other physical supports.

As the three types of users have different functional requirements, the operational processes will been analyzed and illustrated respectively as follows.

Generally speaking, knowledge engineers firstly acquire explicit and tacit knowledge from related experts, enterprise′s original nonstructural database and many other channels with the knowledge acquisition module. Secondly, the core domain concepts are extracted and enterprise ontology database is constructed and maintained by the ontology management module. According to the hierarchical structure of ontology, knowledge engineers annotate the semantic information and definition for the cases, and build case classification index. Finally, the metadata of the defined cases is stored in the case base while non-structured data or the original documents are stored in multimedia library with the XML format.

Normal users follow a different process. Firstly, several key words are input through the case retrieval processing module, and then semantic annotation will be added based on the ontology, the user′s profile and the retrieval history. Then the user′s query can be represented by the semantic vector, which will be matched with source cases later. Finally, cases beyond a certain threshold are sent to the client application layer. Sometimes, failure cannot be avoided. When the users′ needs fail to be met, some cases will be combined together and refined according to ontology by case modify module. In this way, a new case will be formed and stored in case database through case learning module, while other additional information will be added to the multimedia library. It can be seen obviously that ontology management module provides semantic support through the whole process of dynamic management.

System administrator realizes the management of different user permission with reference to the users′ records in the user information database through the system management module.

As business logic processing layer and data service layer play an important role in the dynamic knowledge repository system, they will be discussed further in the part B and part C.

B. Business Logic Processing Layer

The business logic processing layer of the dynamic knowledge repository system put forward in this paper includes the following main function modules.

1) System Management Module：It is responsible for controlling the access authorities of the different types of users. For example, the administrator manages users′ access; knowledge engineers deal with knowledge management and maintenance, while normal users retrieve knowledge.

2) Knowledge Acquisition Module：Two ways can be adopted to acquire domain knowledge. One approach is to arouse the domain experts′ initiatives to obtain knowledge through the brain storm or Delphi method, which is simple, direct and efficient. But certain dependence on the experts will be the main problem. The other way is to dig for knowledge from the present existing results, including the enterprise non-structural database, internal materials, the patent documents, intranet, extranet, BBS and other internal communication, which will be the preparation for the ontology construction.

3) Ontology Management Module：Based on the analysis of the domain scope and characteristics, knowledge engineers extract the core concepts as well as their respective sub-categories from the enterprise. Also, the attributes of each sub-category and the constraints of the attribute are also given. Finally, architecture of ontology will be set up with various interrelationships, such as project classification system, enterprise business department classification system and so on. All these domain ontology will form ontology modeling and be stored in ontology database. The detail will be illustrated in the section IV.

4) Case Definition Module：With reference to the ontology model, some key features are extracted from the documents, the existing project cases and long-term accumulated experiences, which are used to express the corresponding cases in the particular way. Additionally, the ontology- based case classification index mechanism will be established in order to organize the cases clearly.

5) Case Retrieval Processing Module：The function of this module is to segment and analyze the users′ query, and then extend to the different extracted words. The words in user queries can directly map to the concepts, attributes or the instances of ontology, with which every dimension of query vector can be replaced in the semantic view.

6) Case Reasoning Matching Module：The function of this module is to achieve the semantic matching between the user′s query and the annotated cases. Much work is to expand synonym concept and clarify the ambiguous query information or intension. Based on the results of user′s query pre-processing, query vector with semantic expansion is generated for the semantic matching with the representation vectors of source cases in the case database. After computing the similarity between them with certain retrieval strategy and algorithm, cases beyond certain threshold will be returned in order.

7) Case Modify Module：As the scale of the knowledge repository is so limited that it cannot satisfy all users′ demands, the module provide the function of case combination and adjustments according with ontology defined to form new cases.

8) Case Learning Module：The function of this module is to learn the modified cases automatically according to certain rules and enrich the case database gradually. Specific information will be described in section V .

C. Data Services Layer

According to the category of enterprise knowledge, they are stored in the User Information Database, Ontology Database, Case Database and Multimedia Library.

User information database is designed to store all the users' personal background information, retrieval history and reuse records. The prescriptive documents and relational database reflect users′ needs and preferences in order to improve the pragmatic retrieval.

The ontology database is suggested to maintain the domain concepts, the properties, the attribute constraints and relations between these concepts, and finally form the concept model with clear structure. Ontology is so important that many other parts of the knowledge repository system are established based on it. It contributes to the domain knowledge reasoning, case retrieval, matching and learning for dynamic knowledge repository management.

The case database is prepared to store the accumulated cases of the enterprise for a long term. It aims at offering reference for the subsequent ones, which is vital to realize knowledge reuse effectively.

Multimedia library stores semi-structural even non-structured data such as the corresponding document of project cases, related project design, the flow chart, technology and method, the source code, photos, and video conference and so on.

IV. Construction Of Enterprise Ontology

It is obviously that enterprise ontology is the basis of the whole knowledge system and determines the performance as well as the quality of the operation. Therefore, how to establish enterprise ontology correctly, effectively and logically is very important.

Based on the dynamic knowledge repository framework above, we adopt the ″framework method″ [4]

given by Uschold & Gruninger to build the enterprise ontology with ontology modeling tools protégé 3.3.4. Taking an information system consulting company as an example, Fig. 2 and Fig. 3 illustrate the hierarchy and the fragment of the ontology respectively. Firstly, according to the outline, we identify enterprise domain ontology scope in many ways with knowledge engineers and other experts, such as research, interview, brainstorming and so on. Secondly, the core domain concepts are extracted after the analysis and evaluation, and the interrelation and hierarchical structure are defined as well. The main relationships of the ontology may be <is A>, <a part of>, <equal to>, <similar to>, <instance of> and <attribute of>. Fig.2 shows synonymous relationship <equal to>and hyponymy such as <subclass of>. Finally, the attributes of each sub-category and the constraints of the attribute are also given and the ontology is described with OWL.

A small section of the source code with OWL is shown in the following, which is generated automatically by the ontology modeling tools protégé 3.3.4 after the ontology is set up. As we can see, ″Video conference system″ is equal to ″Video session system″, and ″Multi-media project″ is the subclass of ″Project″.

```
</owl:Class>
   <owl:Class rdf:ID="Video_conference_system">
      <owl:equivalentClass
         rdf:resource="#Video_session_system"/>
      <rdfs:subClassOf
         rdf:resource="#Multi-media_project"/>
</owl:Class>
</owl:Class>
  <owl:Class rdf:ID=" Multi-media_project ">
        <rdfs:subClassOf rdf:resource="#Project"/>
</owl:Class>
```

Furthmore ,object properties are defined as follows：

```
<owl:ObjectProperty rdf:ID="has_name"/>
<owl:ObjectProperty rdf:ID="has_budget"/>
<owl:ObjectProperty rdf:ID="has_constructors"/>
<owl:ObjectProperty rdf:ID="has_goals"/>
<owl:ObjectProperty rdf:ID="has_owner"/>
<owl:ObjectProperty rdf:ID="has_requirement"/>
```

As ontology exerts strong influences on the utilization of knowledge repository, the evaluation method based on the feedbacks of users is adopted in the domain ontology assessment. That is to say, knowledge engineers analyze the richness of ontology information (including concepts, attributes, and the definition of an example) and semantic intensity (including ontology structure and relationships) based on the basis of users′ satisfaction.

Except the evaluation, domain ontology maintenance is also necessary. It mainly refers to a series of adjustments, error corrections, perfect and adaptability maintenance work. Correction maintenance focuses on put right in the use. Perfect maintenance refers to ontology expansion work with the increase of knowledge. Adaptive maintenance points to the refreshment of the structure, attributes and relationships with the changes of environment. With tracking and management of ontology model continually, it will provide better support for the dynamic knowledge repository.

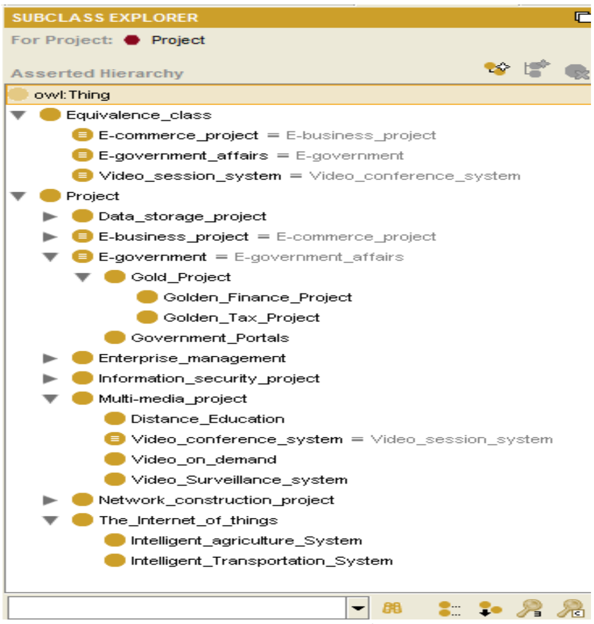

Figure 2. Ontology hierarchy of project types

Figure 3.Ontology fragment of project types

V. Ontology-Based Case Database Construction

A. Case Representation

Actually, case representation is a kind of knowledge expression, which code knowledge into a set of data structure for computer. In this paper the case database is described as: $\{case1, case2, ..., case\ n\} = \{C_1, C_2, ..., C_n\}$; and case C_i is illustrated with ordered pairs <case ID, case title, initial problem description, solutions description, additional information>.Let we take C_i=<ID, TI, IN, SO, AD> for short. The unique identifier of one case is expressed with ID . SO refers to the case solution and the specification, including performance, causes, the main problems, the economic and social benefits, which will be an significant basis of case reuse. AD is related with corresponding documents and multimedia content. Document element and other unstructured data are more suitable to store in the form of multi-media format, which is convenient for the user to understand the whole case.

We divide the case retrieval into case title part represented by TI and initial problem description part represented by IN , which are the basis of the case retrieval. With the consideration of ontology database, feature vector is extracted from these two parts through analysis and expressed by $C_i = \{(t_{i1}, w_1),(t_{i2}, w_2),\ldots,(t_{im}, w_m)\}, i = 1,2,\ldots,n.$ and t_{ij} represents j^{th} feature of i^{th} case, W_{ij} is the corresponding weight of the j^{th} feature, $1 \le j \le m$.

B. Case Organization

In order to improve the speed of case retrieval, the system organizes the cases in the hierarchical structure according to case classification ontology. Then index mechanism is built on the different fields, such as the project classification or industry index, as shown in Fig.4. So, when the user input with some search features, appropriate index is adopted to search the optimized cases efficiently with sematic expansion.

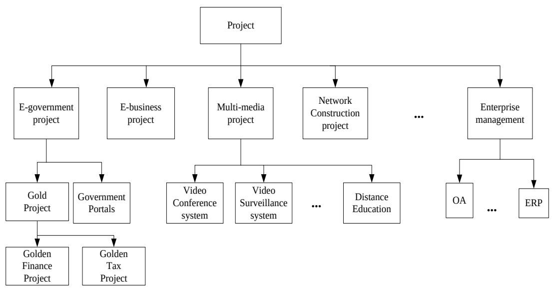

Figure4. The case organization based on project types

C. Case Retrieval

The case retrieval is conducted with the semantic expansion based on the enterprise ontology. The reason the semantic query expansion with ontology is that in query language there are several situations as follows.

Firstly, there are many synonyms. It is quite common in natural language, for example ″E-commerce″ and ″E-business″ both mean commerce conducted electronically. The relationship between these words is called synonyms in the ontology. Besides, sometimes users prefer taking some well-known words for short, for example, ″online-offline″ is often replaced with ″OO″ for short. So, when one user enters some key words, it can be extended by ontology to its synonym.

Secondly, there are many concepts with ambiguities and pragmatic environmental differences. In many cases, the phenomenon of polysemy appears. Take the word ″project″ as an example, it may mean engineering in the broad perspective, or scientific research in the scientific field, while sometimes it means program or planning. Besides, when user input ″finance″ as his application industry demand, it may be considered as the government industry or financial industry. If a record of the user's personal information is e-government consultant and many of historical records are related with the government program rather than financial industry project, we will prefer to define the application industry as government. Thus, in order to eliminate the ambiguity, we should consider user′s background information firstly to clarify its specific and tacit implications

Thirdly, there are some words of superordinate and subordinate concept. In many cases, only through this relationship can we retrieve the potential information, such as project, multimedia and video conference.

According to the case representation method illustrated in the part A of this section, suppose there are m keywords, and one semantic vector can be expressed as $T_i = [t_{i1}, t_{i2} \cdots t_{im}]^T \; i = 1,2,\ldots,n$.

Based on semantic expansion with the domain ontology, each of the key word can be expressed as a semantic vector, which will be used to match with the keywords defined in the source cases. According to the description in the part A of this section, as the importance between case title part and initial problem description part will be different in concrete applications, we should consider the weight of one keyword appeared in the title part and initial problem description part. Let α represent the weight of the title part; let β represent the weight of initial problem description part and $\alpha + \beta = 1$. Generally speaking, the main content will be shown more obviously in the title of one case, so we define $\alpha > \beta$ in the case retrieval process of this paper.

Suppose C as a candidate case set, and $C_i \in C$ is one candidate case of the candidate case set, then C^* represents the query vector that is consistent with the user′s demand. We define the feature vector of a case expressed as $T_i = [t_{i1}, t_{i2} \cdots t_{im}]^T$. So the semantic similarity between C_i and C^* can be defined as (1).

$$SIM(C_i, C^*) = \alpha \cdot SIM_{TI}(C_i, C^*) + \beta \cdot SIM_{IN}(C_i, C^*) \quad (\alpha > \beta, and, \alpha + \beta = 1) \tag{1}$$

In the formula above $SIM_{TI}(C_i, C^*)$ means the similarity of the title part and $SIM_{IN}(C_i, C^*)$ means similarity of the initial problem description part.

While computing the value of $SIM_{TI}(C_i, C^*)$, the frequency of each keyword in the vector space in every case title is ignored. Therefore, if j^{th} keyword doesn't exist in the title of the case C_i, in another word, j^{th} keyword isn't important in the title of the case C_i, and then $t_{ij} = 0$. Else, j^{th} keyword is very important in the title of the case C_i, then $t_{ij} = 1$.

However, while computing the value of $SIM_{IN}(C_i, C^*)$, the frequency of each keyword in the vector space in initial problem description part of every case should be considered, as the descriptions of cases are more complicated and can reflect much information. So, the frequency of j^{th} keyword will be numbered. In this part, let t_{ij} $(i = 1,2,\cdots,n\,;\, j = 1,2,\ldots m)$ denote the frequency of j^{th} keyword in the description of i^{th} case and m is the total number of the terms. So in the same m dimension vector space the description of i^{th} case can be represented by $T_i = [t_{i1}, t_{i2} \cdots t_{im}]^T$. Then we can realize that value of t_{ij} denotes the importance of j^{th} keyword.

As $t_{ij} \ge 1$, the following step is to regress the feature vector of the case description to [0, 1] for convenient calculation. Let

$$d_{ij} = \frac{t_{ij} - \min_{1\le j\le m}\{t_{ij}\}}{\max_{1\le j\le m}\{t_{ij}\} - \min_{1\le j\le m}\{t_{ij}\}} \quad (2)$$

In this way, not only the title part but also the case description part can be represented in the following expression $D_i = [d_{i1}, d_{i2} \cdots d_{im}]^T$. Particularly in the title part as the value of $t_{ij} \in [0,1]$, then $D_i = [d_{i1}, d_{i2} \cdots d_{im}]^T$ is the same as $T_i = [t_{i1}, t_{i2} \cdots t_{im}]^T$.

Set the user′s query vector as standard, and then the query vector space should be $C^* = W^T \bullet I = (w_1, w_2, ... w_j, ..., w_m)$ （I represents a unit matrix）. W^T is set with the scoring method. Only the user can understand his demand most clearly, therefore, W_j^T is calculated by the keywords preferences of the user. We divide the user′s keyword preference level into the following fuzzy set:{very important, important, common, unimportant, very unimportant}. For quantitative analysis, the fuzzy set can be mapped into the vector{5，4，3，2，1}. If the user′s preference vector for the m keyword is $(x_1, x_2, ..., x_m)$ $1 < x_j < 5$, and integer, then

$$w_j = \frac{x_j}{\sum_{j=1}^{m} x_j} \quad (3)$$

Therefore, the similarity between the query information and source case information in different part is calculated with （4）and（5）respectively.

$$SIM_{TI}(C_i, C^*) = \sum_{j=1}^{m} w_j \cdot \cos(T_i, C^*) \quad (4)$$

$$SIM_{IN}(C_i, C^*) = \sum_{j=1}^{m} w_j \cdot \mathrm{COS}(D_i, C^*)$$
$$= \sum_{j=1}^{m} w_j \cdot \frac{D_i \cdot C^*}{|D_i| \cdot |C^*|} \quad (5)$$
$$= \sum_{j=1}^{m} w_j \cdot \frac{\sum_{j=1}^{m} d_{ij} \cdot c_j^*}{\sqrt{\sum_{j=1}^{m} d_{ij}^2} \cdot \sqrt{\sum_{j=1}^{m} c_j^{*2}}}$$

While computing the similarity of query vector C^* and semantic vector C_0, C_1, $\cdots C_n$, remember to represent the largest similarity as $SIM_{\max}(C_i, C^*)$. If the value is larger than or equal to the given threshold ε, the case will be added into the search results. Finally, the search results sort according to the similarity and be sent to the user.

D. Case Reuse and Learning

Generally speaking, the scale of the initial case base is so limited that cannot satisfy all the different needs of customers, so the dynamic management mechanism becomes urgent. It mainly includes the realization of the case reuse, modification and case learning on the basis of case retrieval. Fig. 5 shows the flow chart of this dynamic mechanism.

In the process of case reuse, two main problems should be considered. One is the differences between the new problem and the retrieval results. The other is which part of the result can be reused. We define those cases met the user′s demand group in the matching set expressed by G, that is $G = \{C_i | SIM(C_i, C^*) > \varepsilon\}$.

If $G = \Phi$, then decompose the query vector by level in accordance with ontology and calculate the similarity for searching the matching set which is not Φ. Mark this level as starting point, and combine the cases from down to up with the ontology rules which can coordinate the constraint and standard between different cases in the process of combination. Two methods of combination are appropriate: exhaustion and genetic algorithm. When the number of case combination choices is small, that means all possible combinations can be listed and the feasible solution can be found out. We consider the one who got the largest similarity with the new case as the optimal solution. Otherwise, genetic algorithm will be more efficient. Then, the process of case adjustment begins.

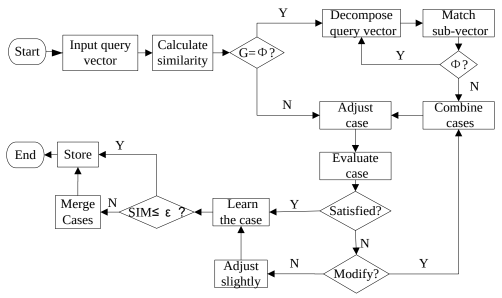

Figure5. Flow chart of dynamic management

If $G \ne \Phi$, the process of case combination skips and directly gets into the case adjustment process. As case adjustment will also need to consider some other information besides the case similarity, it should be adjusted and decided according to case description, solutions and other affiliated information with the help of persons, in the manual or semi-automatic way. We define the case after adjustment as target case, which is more close to new problem than the source case.

The case is evaluated in practice, and then a new case which meets our requirements will be learned. In order to avoid the redundant information in the knowledge repository, similarity between the new case and the source case will be also calculated to determine the new case will be stored or be processed further. Set certain threshold ε, S represents source case while N is new case, so if $SIM(S, N) \le \varepsilon$, store the new case; else, merge them together.

In this process, the ontology provides the semantic support for the case retrieval, matching, combination, adjustment and case learning.

E. Validation and Analysis

In order to verify the retrieval effectiveness, according to the research work in this paper, a simple prototype system of consulting industry is developed based on SQL database and C # language.

It’s difficult to make the case database rich at the very beginning. Therefore, we mainly collect relevant data from the enterprise original relational database system, Intranet, and project materials and then sort into 40 information system project consulting cases with analysis. These cases are deposited into the knowledge repository in the description and organization way illustrated above, which can be seen as the preparation for the operation.

Suppose that a consultant who is mainly responsible for e-government projects needs to retrieve some information of history projects which provide project management services for video conferencing system in the finance field. Besides, if Ministry of Finance is main project stakeholders, that will be better. So, the key retrieval vector is extracted as Industry, Project Types, Services and Main Project Stakeholders. This consultant enters the following vector value:

$$T_{ij} = [\mathit{Finance}, \mathit{Video\ conferencing\ system}, \mathrm{Project}\ \mathit{Management}, \mathit{Ministry\ of\ Finance}]^T$$

The case retrieval interface is shown in Fig. 6.

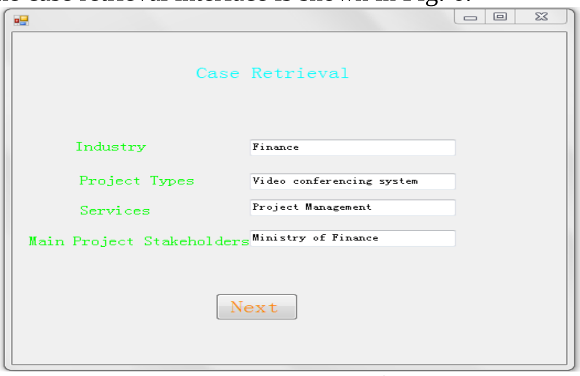

Figure 6 User query interface

Also, his preferences for these keywords are defined successively as important, very important, very important, and unimportant respectively, which is shown in Fig.7. In this way, the score could be described as $(\mathbf{4,5,5,2})$ and got W^T=[1/4,5/16,5/16,1/8]^T with (3). According to the domain ontology above and the user′s information, the retrieval vector can be expressed more precisely with semantic and pragmatic expansion, which is shown in detail in the Table. I. Particularly, the dimension of finance is expressed as government rather than financial industry, because the information of the consultant shows that he is responsible for the E-government services and the major project stakeholder is Ministry of Finance.

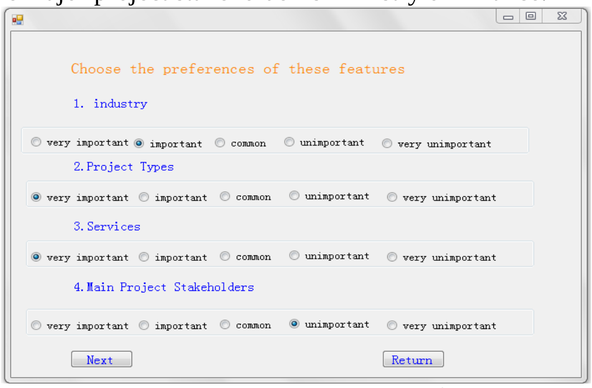

Figure 7 Weights determination interface

In this paper, we set $\alpha = \mathbf{0.6}, \beta = \mathbf{0.4}$; the threshold value of similarity $\varepsilon = \mathbf{0.5}$, then the consultant′s retrieval vector was: $C^* = W^T \bullet I = (\mathbf{1/4, 5/16, 5/16, 1/8})^T$. Computing with the retrieval algorithm and the expansion vector shown above, cases those meet the formula $SIM(C_i, C^*) \geq \varepsilon$ are delivered as Fig.8.

TABLE I.
RETRIEVAL FEATURE VECTOR AND SEMANTIC EXPAND INFORMATION

t_{ij}	x_j	w_j	Vector Value	synonymy	Pragmatics	Hyponymy
Industry	4	1/4	Finance	——	Government(Y); Financial Industry(N)	Government
Project Types	5	5/16	Video conferencing system	Video session system	——	Multi-media
Services	5	5/16	Project Management	PM; Supervisor	——	Management
Main Project Stakeholders	2	1/8	Ministry of Finance	Treasury Department; Finance Bureau	——	Person; Organization

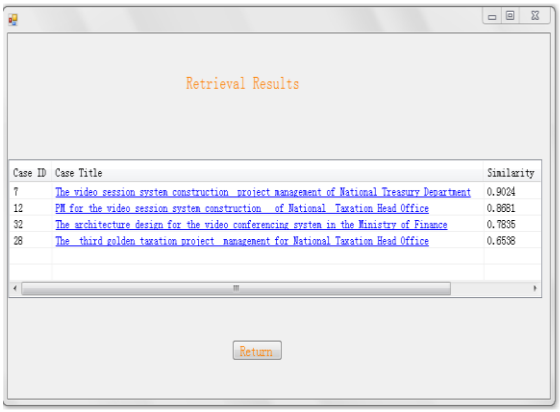

Figure8. The matching cases and similarities

Take the optimal matching case 7 as example. The process is described as follows.

$$T_{TI} = (\mathbf{1/4, 5/16, 5/16, 1/8})^T$$

$$SIM_{TI}(C_7, C^*) = 1$$

$$T_{IN} = [\mathbf{1, 4, 10, 1}]^T$$

$$D_7 = [\mathbf{0, 1/3, 1, 0}]^T$$

$$SIM(C_7, C^*) = 0.7559$$

$$SIM(C_7, C^*) = 0.6 \times 1 + 0.4 \times 0.7559 = 0.9024$$

The traditional way of information retrieval is based on keywords that the computer cannot understand user′s potential semantic and personalized query demand. Therefore, we conduct the semantic and pragmatic retrieval research based on the domain ontology and user's information. After the analysis, semantic and pragmatic retrieval can get the best precision and recall rate of the matching result among traditional retrieval method, sematic and pragmatic method, as is shown in TABLE II.

TABLE II.
THE COMPARISON OF THE RETRIEVAL METHOD

	Retrieval Results	Retrieval Matching Cases	Actual Case Number	Precision	Recall
Traditional searching	3	2	8	0.6667	0.375
Semantic retrieval	5	4	8	0.8	0.5
Semantic and pragmatic retrieval	8	7	8	0.875	0.875

VI. CONCLUSION

The dynamic knowledge repository construction is a hot issue in the field of knowledge engineering at present. This paper focuses on the construction method of knowledge repository based on ontology and CBR. An integrated framework is put forward and the operation processes of dynamic knowledge repository system are analyzed. With the construction of ontology the query of the user can be expanded with more semantic information and the case database can be formed more precisely. A retrieval algorithm is designed and the dynamic mechanism of the knowledge management is illustrated. Moreover, taking a consulting company as the case study background, the retrieval process is verified, which proves the high accuracy and completeness of the retrieval algorithm.

As the initial case samples are small and centralized, further work should be done to conduct quantitative analysis based on a large size and wide distribution of case samples.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 71102111 and Beijing Institute of Technology under Grant 3210050320908.

REFERENCES

[1] Annette M. Mills, Trevor A. Smith. Knowledge management and organizational performance: a decomposed view [J]. Journal of Knowledge Management, 2011, Vol.15 Iss1, pp.156–171.
[2] K. D. Joshi, Lei Chi, Avimanyu Datta, Shu Han. Changing the Competitive Landscape: Continuous Innovation Through IT-Enabled Knowledge Capabilities [J].Information Systems Research, Vol.21, No.3, September 2010, pp.472-495.
[3] Chou, Shih-Wei. Knowledge creation: absorptive capacity, organizational mechanisms, and knowledge storage/retrieval capabilities [J].Journal of Information Science, December 2005, vol.31, pp. 453-465.
[4] Stanford University Knowledge System Laboratory. A translation approach to portable ontology specifications [R]. Stanford: Stanford University Knowledge System Laboratory, 1992.
[5] Norbert Gronau, Frank Laskowski .Using Case- Based Reasoning to Improve Information Retrieval in Knowledge Management Systems [J].Springer-Verlag GmbH, 2003; 2663, pp.94-102.
[6] Wimalasuriya, Daya C., Dejing Dou. Ontology-based information extraction: An introduction and a survey of current approaches [J].Journal of Information Science, June 2010, vol.36, pp.306-323.
[7] Wenhuan Lu, Ikeda, Mitsuru.A uniform conceptual model for knowledge management of international copyright law[J].Journal of Information Science, February 2008,vol.34,pp.93-109.
[8] Akbari, Ismail, Fathian, Mohammad. A novel algorithm for ontology matching [J].Journal of Information Science, June 2010, vol.36:pp.324-334.
[9] Chimay J. Anumba, Raja R.A. Issa , Jiayi Pan, Ivan Mutis.Ontology-based information and knowledge management in construction[J].Journal of Knowledge Management,2008, Vol.8,Iss3, pp.218–239.
[10] Liao Liangcai, Qin Wei, Shu Yu. Ontology-based Dynamic Knowledge Management System [J].Computer Engineering, 2009, 35（16）, pp.256-261.
[11] Gao Huiying, Yan Zhijun. CBR based Multi-agent System Model for Case Retrieval of Information Systems [J].Computer Engineering and Design, 2008, 29(5), pp.1226-1228.
[12] Gao Huiying, Zhao Jinghua. Ontology-based Enterprise Content Retrieval Method [J].Journal of Computers, 2010, 5(2), pp.314-321.
[13] Huiying Gao, Qian Zhu. Semantic Web based Multi-agent Model for the Web Service Retrieval. Proceedings of International Symposium on Computer Network and Multimedia Technology, 2009.12, pp.897-900.

Huiying Gao, Dr. Associate professor, was born in Shandong Province, China, in 1976. She received her doctor degree in management science and engineering from Beijing Institute of Technology in 2003. Now she is an associate professor in the school of Management and Economics, Beijing Institute of Technology. She has ever been in Technical University of Berlin, Germany to do her Ph.D. research work from 2002 to 2003. From 2008 to 2009 she has visited Karlsruhe Institute of Technology, Germany for half a year. Her current research interests include theory and method of information systems, content and knowledge management, semantic retrieval, intelligent information system etc. Ph: +86 (10) 68918830.

Xiuxiu Chen was born in Shandong Province, China, 1988. She is a Ph.D. degree candidate in the school of Management and Economics, Beijing Institute of Technology. Her research work includes information management and knowledge management. Ph: +86 (10) 68918830.