TELKOMNIKA, Vol.11, No.11, November 2013, pp. 6591~6599
e-ISSN: 2087-278X

Research of Semantic Network Knowledge Representation and Query Algorithm based on Relational Model

Li Yue-xin*, Hong Zong-xiang
Faculty of Mathematics & Computer Science of Hubei University
Wuhan of Hubei Province, China, 430062
*Corresponding author, e-mail: 51884285@qq.com

Abstract

Because of the solid foundation, mature technology, rich products of relational database, it has been the first consideration for studying and storing the massive data. How to build an effective knowledge base based on the powerful storage and processing capability is always the key issue for the research and application of knowledge engineering and intelligent system, but also the bottleneck knowledge management turns into the application of technology. In this paper, a relational model of semantic network has been given and discusses with examples the storage of relational data table based on the semantic network knowledge representation. Once semantic network relational model established, the knowledge base that based on semantic network will be set up. In this paper, Semantic network knowledge base query algorithm is studied.

Keyword: *semantic network, relational database, relational model, knowledge base, query algorithm, relational tree, implicit knowledge*

1. Introduction

Semantic Network was the first proposed by J. R. Quillian in his doctoral thesis ,which as an explicit psychological model in human associative memory [1]. Then, it was used as knowledge representation in TLC (Teacher Language Comprehended) designed by J.R. Quillian as well as used for natural language understanding system by Simon in 1972. Currently, Semantic Network was widely used in the field of artificial intelligence and knowledge engineering for its strong and flexible capability to express.

Currently, the tools of storing and using of Semantic Network in domestic and international, mostly based on the ontology's knowledge system, such as Sesame, Jena, 3store, SquishQL other types of storage management system. JENA language based on RDF (Resource Description Framework) API as the core and can be used to achieve the Semantic Network [2]. JENA RDF support the creation, manipulation and query and many other functions, also supports many different data storage technologies. SquishQL language can be seen from the name associated with the SQL language. This is a SQL-based RDF query language, the use of relatively simple, and there are two constraint classes can be used to express patterns and filtering, you can directly use the Web as a database [3]. By definition, ontology is a theory-based knowledge system; research in the theory is widely used to describe the domain knowledge [4]. Therefore, the ontology should be based on the theory, according to actual needs, planning appropriate programs and strategies of knowledge management [5]. Authors are studying the knowledge system of ontology, semantic network representation of specific areas of knowledge.

Based on relational data model and relational database management system's powerful storage, processing capacity and good reliability, in this paper introduce the method of semantic network knowledge representation and technical methods of the storage, use and maintenance in the support of relational databases as well as discuss and analyze the access efficiency under this technology.

Received March 10, 2013; Revised June 21, 2013; Accepted July 22, 2013

2. The Defintion of Semantic Network

Semantic Network uses a way to represent knowledge structure, which mainly combine the concepts and semantic relations with a graph. Just from the graph theory point of view, it is actually a 'direction of the map. Below is a simple semantic network, represented by triples:

Semantic (node1, arc, node2) (1)

Node1、node2 represent the things, concepts and an event of the domain knowledge, arc represents the relations among things, concepts and events. Basic semantic elements are shown in the Figure1.

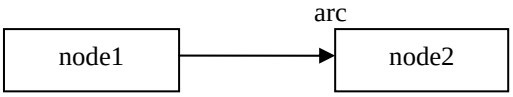

Figure 1. Basic Semantic Elements

Node i is shown with rectangular, arc is shown with directed straight line. Main relations contain:

(1) subordinate relation-- is a ;
(2) assemble relation---- a part of;
(3) classification relation---- a kind of;
(4) inference relation---- derived from;
(5) action or behavior relation---- action or take part in;
(6) time, space relation---- before、after、behind、in、at.

Semantic network system is a knowledge system, which contains a multi-node directed graph. The following is the definition:

G= (V, E) (2)

V stands for the node set in the knowledge system, E stands for the relation set among nodes.

Uncertainty representation of semantic network contains three categories:

1. Uncertainty of semantic relation;
2. Uncertainty of nodes;
3. Uncertainty of semantic network's structure.

In this paper, the uncertainty of semantic relation is the topic, so a topple is defined:

A =(G, F) (3)

G is the directed graph in the (2) sentence, V is the node set in the semantic network: V = (v1, v2, …, vn)

If E(vi,vj) exists, E is the edge in the G, then define:

F= (μi/(vi,vj)) 0 ≤ μi ≤ 1 , i=1,2,…, |E| (4)

μi is the membership of the （vi,vj）, it means the ambiguity of the node (vi,vj) relation; F is a fuzzy relationship of the V×V [6].

3. The Relational Schema and Data Model of The Semantic Network

Two relational schemas can describe the relation between the Start_node and End_node of the directed edge of the semantic network:

Start_Node(NodeID, NodeName, NodeArcIndex, [membership])
End_Node(NodeID, NodeName, ArcAttribute)

Start_Node relation describes the start_node of the directed edge, its attributions are NodeID, means nodeID and NodeName, means nodeName. [content] means that content is optional. Membership is omitted. End_Node relation describes the end_node of the directed edge, its attributions are NodeID, means the nodeID of the end of directed edge, ArcAttribute, and means the relation of node, Membership, means the degree of membership of relation. If abstract it with mathematical language, the two relations constitute a two-dimensional matrix, Start_Node is the row element, End_Node is the column element, ArcAttribute is the value of element.

Described as follows on the fact of Bob and Mary [7]:

Bob and Mary are Grade 6 students of the XX primary school, they are neighbors of the Friendship Community. Bob raises a dog, named peter, Bob's mother at 7 o'clock every afternoon walking the dog in the Community, she can see many the elderly doing exercise.

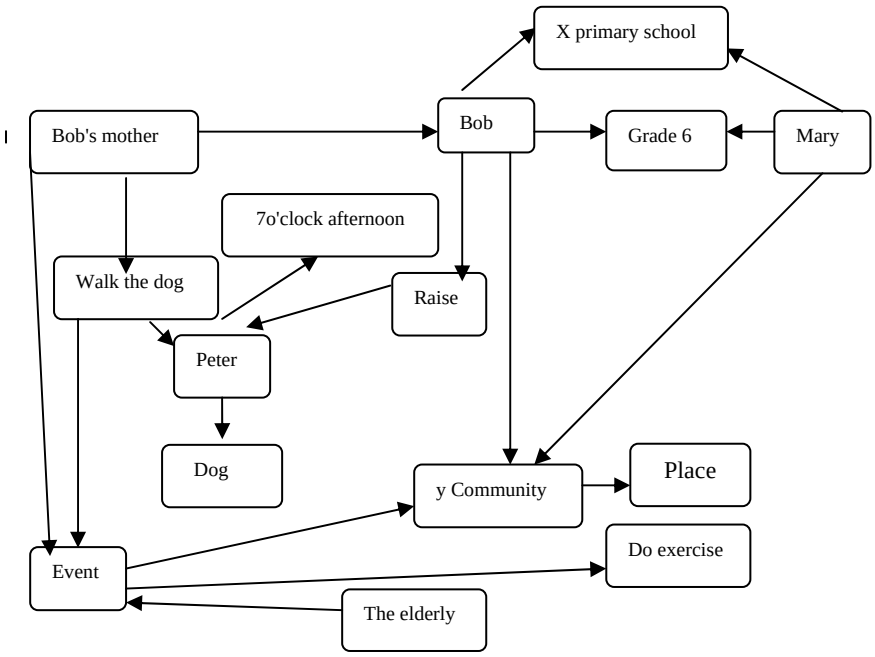

Figure 2. Bob and Mary's Facts are Described with the Semantic Network

The facts can abstract some concepts including (contain the time concept): Bob, Mary, XX primary school, Grade 6、neighbor、dog、peter、Bob's mother、the elderly、Friendship Community、7 o'clock afternoon and also can abstract some events or activities including: Raise、walk the dog、do exercise and so on. So about Bob and Mary's facts can be described with the semantic network in Figure 2 and database relationship table in Figure 3.

Now analyze the space consumption of resources based on the semantic network knowledge representation of the relational model. If there are n nodes in the graph G= (V, E), which is represented by semantic network of the field of knowledge, Start_Node_Table maximum number of table records is n. In the worst case, each node i has the directed edges point to all nodes and End_Node_Table i table number is n, while the number of records in the table is n. Therefore, when a rational database based on the semantic network of the field knowledge representation is implemented successfully, the records numbers are: S (n) =O (n2). Although: S (n) is polynomial, but still very huge. So, how to optimize the database query for improving the efficiency in the use of knowledge will be the focus of follow-up study.

Now analyze the space consumption of resources based on the semantic network knowledge representation of the relational model. If there are n nodes in the graph G= (V, E), which is represented by semantic network of the field of knowledge, Start_Node_Table maximum number of table records is n. In the worst case, each node i has the directed edges point to all nodes and End_Node_Table i table number is n, while the number of records in the

table is n. Therefore, when a rational database based on the semantic network of the field knowledge representation is implemented successfully, the records numbers are: S (n) =O (n2). Although: S (n) is polynomial, but still very huge. So, how to optimize the database query for improving the efficiency in the use of knowledge will be the focus of follow-up study [8-10].

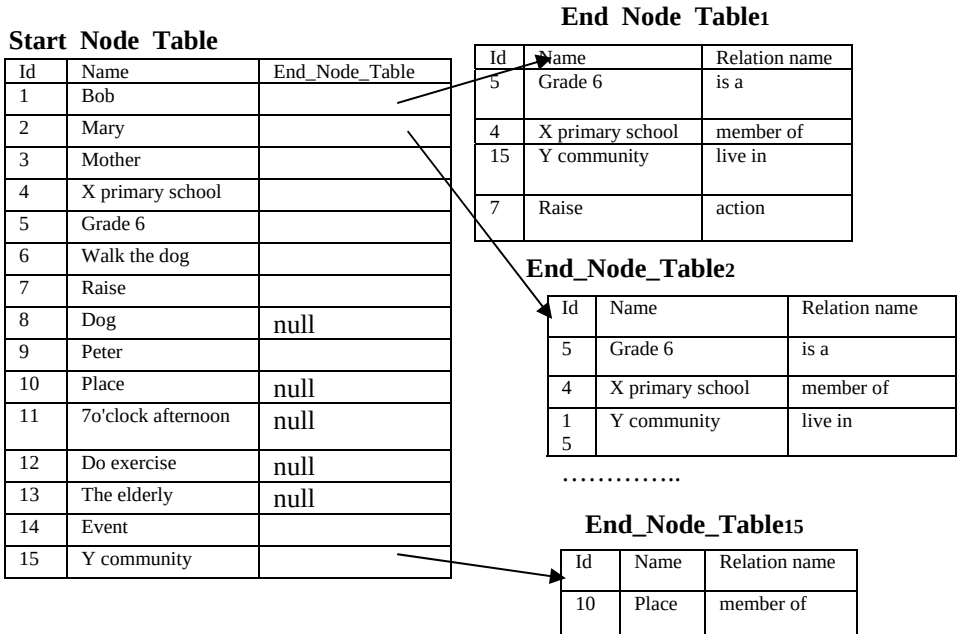

Figure 3. Bob and Mary's Facts are Described with the Database Relationship Table

4. Knowledge Query Alogorithm

a. Introduction

Since the field of AI has been proposed in 1956, the majorities of research workers have study it in an extensive and depth way, but also achieved unprecedented results. The "Knowledge Engineering" concept has been proposed until 1970s and it can be described as a milestone in the AI history. With the emergence of expert system and commercial gradually, the theory and technology of knowledge engineering is increasingly updated. Knowledge automatic acquisition, the theory and technology of knowledge base system and distributed knowledge base system are becoming the main direction [11]. Knowledge representation storage and reasoning queries in the knowledge base is becoming hot point at home and abroad. Natural language query is the most important part of it. Natural language query problem is to study how to make a computer can understand and produce language used by people everyday, and using the natural language for feedback [12]. Authors transfer the relational model of semantic network knowledge presentation proposed in literature1 into a relational tree model and propose forward, backward directed knowledge query algorithm and implicit knowledge query algorithm for the knowledge can not be queried by the directed knowledge query algorithm.

b. Natural language description

The natural language problem is outstanding in early stage and because of concept of knowledge engineering launched, making natural language query problem based on knowledge base system is particularly important. Usually, user input natural language into query interface and the system implements the processing operation in the background, and then put the inquiry results returned to the user. The basic idea of the background processing is: first, splitting words from nature language, pre-treatment, then the semantic analysis, knowledge extraction, template matching and feedback results etc [4]. Among of them, splitting words process is also key words extraction process, so the nature of most natural language query is keywords query. This paper is mainly to study the knowledge query algorithm in the knowledge

extraction module. Knowledge extraction basic idea is: The first step: Representing the knowledge in the knowledge base with

Semantic network, using the relation between start node and end node of directed edge to build a relation model and then will relation model 90 degrees clockwise get relation tree model. The second step: querying in this relation tree model and calling query algorithm, return current knowledge to match template. The third step: return to query results. Among them, the second step is the core in this paper. The following is the simple instructions about the process from semantic network to the relation model and then to the relation tree model:

Described as follows on the fact of Bob and Mary:

Bob and Mary are Grade 6 students of the x school, they are neighbors of the y Community.

The process as following:

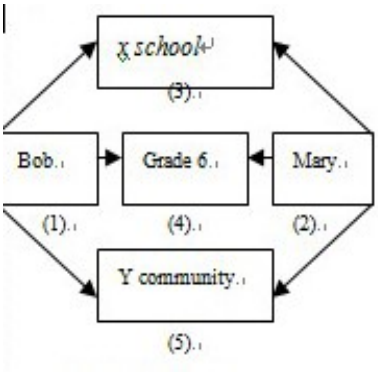

Figure 4. Semantic Network

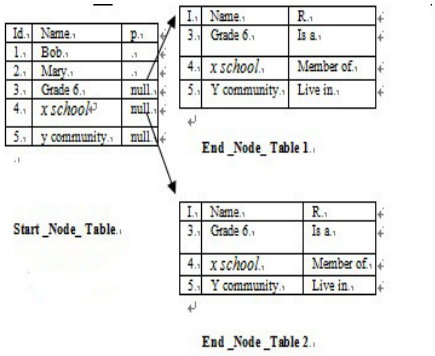

Figure 5. Relation Model

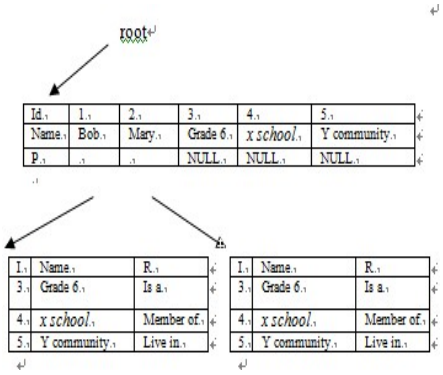

Figure 5. Relation Tree Model

Among, the definition of Start _Node_ Table and End _Node _Table as following:

Start _Node _Table (NodeID, NodeName, EndNodePointer)

End_ Node_ Table (NodeID, NodeName, NodeRelation)

Start_Node_Table describes the start node of directed edge of the semantic network, its attribution are NodeID, means node ID, NodeName means node name, EndNodePointer is a pointer point to the End_ Node_ Table; End_Node_Table describes the end node of directed edge of the semantic network, its attribution are NodeID, means node ID, NodeName means node name, NodeRelation is the relation between the start node and end node (including, a, member of, live in etc).

5. Query Alogrithm Description

a. Sentence template matching algorithm

In this paper, the authors construct a dictionary with knowledge nodes in knowledge base and the words in the dictionary are classificated according to the part of speech to build

table such as nouns table, the question word table, etc. List common sentences constitute sentence template table. User queries from the semantics are divided into two classes: first, for intellectual inquiry, its purpose is to extract knowledge from knowledge base system for the unknown knowledge, such as: "what is the artificial intelligence?", "what is network protocol?" etc, the second is the verification query, that means users have some areas of knowledge, its purpose to through the knowledge base system for these knowledge or supplement verification, such as "Is computer system constituted by the hardware and software system?" [13], etc. Using concept-attribute model collect sentence templates: using C stand for concept and if there are multiple concepts with C1, C2 and C3..., using A stand for attribute, if there are multiple attributes with A1, A2, A3.... Using A.V stand for value of attribute A, use point signal stand for the membership of attribute A, such as the attribute A of concept C can be stand for C.A , and its value for C.A.V. Therefore, inquiry means the query of concept query, concept relationship and relationship among concepts query, part of the template as follows [14] :

< Template 1>：：what are the attributes of? (query concept description)

< Template 2>：：what is the description of C? (query attribute description)

< Template 3>：：what is the relation between C and C1? (query relation among concepts)

For the convenience of discuss, set the Tm stand for word biggest segmentation threshold, T for similarity threshold, Ts for match successful threshold, Wd for matching degree, the sentence pattern template matching algorithm (SM) fake code as follows:

SM algorithm is as follows:

Step 1: For each phrase i set a similarity threshold T with fuzzy set;

Step 2: According to the structure of the dictionary, from the start of natural language with the maximum matching splitting and then searching in the current Chinese characters of the string , estimating the string matching similarity of Ti and word segmentation threshold Tmi, if Tmi≦Tm and Ti≧T, then word matching success; Otherwise, call multilevel similar word library, getting the content string under the best approximation matching conditions.

Step 3: Transferring sentence pattern structure information into a query semantic information and using fuzzy set to match the template , if Md ≧Ts, the answer is matching success, or turn to Step2;

b. Direct knowledge query algorithm

Relation tree is not trees, but combined with the adjacency lists and relation table two storage structures, its query will contain the two storage structure and tree structure characteristics. The upper is adjacency structure, the under is relation table. If semantic network has N nodes, each node has index ID number. Direct knowledge is the knowledge, which can be queried in the relation mdel with the description of the fact 's key words. Based on this query, it is divided into two classes: forward and backward knowledge query. Forward query namely from top to bottom to queering and the keywords extracted from natural language have end nodes. If keywords have no end nodes in the query process and need back, this can be called backward query. In this paper, select c++ as the algorithm language [15], query algorithm is as follows:

(1) Forward knowledge Query Algorithm (FKQA):

```
Void ForwardKQ (kT &G)
{
int i, j, k, Nodenum, count1=1, count2=1,*b;
int ID[Nodenum]   // define the node Id array
cin>>Id>>Nodenum    //input the node would be queried and node number
for (i=1;i<=G.Nodenum;++i)
cin>>ID[i];  //input the all nodes Id number
while (Id! =ID[j]&&j<=Nodenum)
{
++j;
++p;
count1++;
}  //query adjacency table
if(j>=Nodenum) return false;
else cin>>Id; b=p;
while(Id! =ID[k]&&k<=Nodenum)
```

```
{
++k;
++b;
count2++;
}  //query relation table
cout<<p.ID[count1].Name<< b.ID[count2].R
<<b.ID[count2].Name;
} // Forward knowledge Query Algorithm (FKQA):
```

(2) Backward knowledge Query Algorithm (BKQA)
c. Implicit knowledge query algorithm

```
void BackwardDKQ(KT &G)
{
int i, j, ID[Nodenum],j=0,count=0;
int *p[],*q;
for(i=1;i<=G.Nodenum;i++)
 cin>>ID[i]; // input the all nodes Id number
cin>>Id>>Nodenum; // input the node would be queried and node number
p[0]=root;
while(j<=Nodenum)
{
j++;
if(p[j]=null)  cout<<p[j].Id;
while (Id! = (p[j].Id)&&k<=Nodenum)
{
++k;
 ++q;
count1++;
} // query relation table
count2++;
}  // query adjacency table
cout<<p[j].ID[coun2].Name<<q.ID[count1].R
<< q.ID[count1].Name;
}// Backward knowledge Query Algorithm (BKQA)
```

Direct Knowledge in knowledge base can be directly queried by the above algorithms, but some knowledge, implied and can't be queried directly, are called implicit knowledge Implicit knowledge can be roughly divided into two types [16]. One kind is special value vocabulary knowledge, for example: "student", usually in the database have no "student" this abstract vocabulary, instead concrete vocabulary such as "pupils", "high school students", "college students" etc; the another kind is the relevant operation vocabulary knowledge, it is caused by the correlation among knowledge, for example: "real income", it equals income minus the cost income, but usually not in knowledge base in such vocabulary (easy to produce data redundant). Therefore, the implicit knowledge has a common feature of "abstract", and an effective solution is specification, namely transfer implicit knowledge into direct knowledge. According to the particularity of implicit knowledge, we can stand foe implicit knowledge with concept-tree. The "student" concept-tree is as follows:

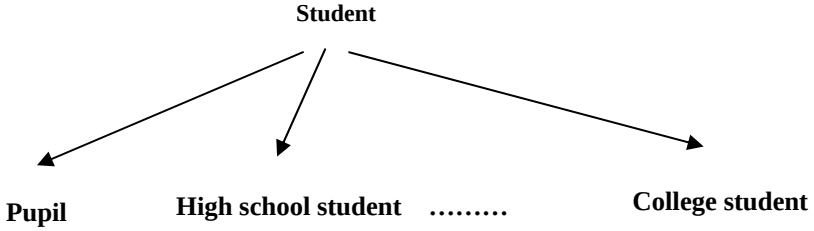

Figure 6. "Student" Concept-tree

In the Figure transferring the student into specific vocabulary such as the pupil, high school student, college student, so the first step of implicit knowledge query algorithm is transferring, the next step is direct knowledge query algorithm, algorithm is as follows:
Implicit Query Knowledge Algorithm (IQKA):
Step 1: Splitting words from the natural language and queering whether or not contain implicit words, if there are implicit words, turn to Step 2, if not, using direct knowledge query algorithm;
Step 2: Querying the concept-tree of implicit knowledge and representing the specific words and relations among them, then check rules of grammar library;
Step 3: Output query results, matching with the current template, and return results to users.
d. Algorithm analysis

In this paper, mainly analyze the direct knowledge query algorithm. Forward direct knowledge query algorithm's basic idea is: through the level traversal query methods from top to bottom, return the key words and relations among them. For example: user ask a question "Is Bob's address y community?' Results returned: Bob Live in y community, the subsequent steps are template matching, feedback to the user's query results; this algorithm can effectively deal with direct knowledge query [17]. The complexity of the algorithm is O (n2), is a polynomial level, when n is much bigger, and query time is longer, so propose the optimizing algorithm is the research key [18].

6. Conclusion

This paper presents a relational model of the semantic network knowledge representation. Through an example, given the example of relational schema and corresponding database relational tables, verify the Semantic network knowledge representation and the contact of the relational schema. This approach can be seen with a good solution to solve a problem, which uses a two-dimensional rational table point to all nodes to represent multi-dimensional non-structural knowledge. Paper due to limited space, do not address how to query the database (how to use the knowledge), namely the use of the fact that reasoning to get the facts related to known knowledge and how to get the facts implied in the known facts. In addition, there are a few of defects in the semantic network knowledge representation itself, such as the methods represent knowledge are not uniform and it is more difficult to represent a complete, large-scale knowledge system. These problems are ongoing in-depth study and the study results will be published in another paper.

In this paper, authors propose relation tree model on the base of relation model and the forward, backward direct knowledge query and implicit knowledge query algorithm. Those algorithms effectively solve the query problem in knowledge base. Because of the limited space paper, there are some problems exist ,such as a small amount of templates, writing the sentence template matching algorithm and implicit knowledge query algorithm in fake code, and some defects of relation tree model itself and query algorithm is polynomial level etc, put forward more robust optimization algorithm is the focus of future research r, the authors have considered parallel computing theory is introduced to come in to solve those issues, the research results of the related problems will be published in subsequent papers .

Acknowledgement

Fund Subsidized: The Hubei Foreign technical cooperation project. No.2012IHA014

References

[1] Lu jianjiang, Zhang yafei, Miao zhuang, Zhou bo. Principles and technology of semantic network. Science Press, Beijing. 2007.
[2] Deng zhihong, Tang shiwei, Zhang min. Ontology Research. *Journal of Peking University (Natural Science Edition).* 2002; 38(5): 730-738.
[3] Neches R, Fikes R Finin T, et al. Enabling Technology for Knowledge Sharing. *AI Magazine*. 1991; 12(3): 36-56.
[4] Pan Z, Zhang X, Heflin J. DLDB2: A Scalable Multi-Perspective Semantic Web Repository. In W1 08: Proceedings of the International Conference on Web Intelligence. *IEEE Computer Society Press*. 2008.

[5] V Haarslev, R Moiler. *RACER System Description.* Proc. of the International Joint Conference on Automated Reasoning. Heidelberg: Springer-Overflag. 2002; 701. N705.
[6] Adnan Tawafan, Marizan Bin Sulaiman, Zulkifilie Bin Ibrahim. Adaptive Neural Subtractive Clustering Fuzzy Inference System for the Detection of High Impedance Fault on Distribution Power System. *IAES International Journal of Artificial Intelligence*. 2012; 1(2): 204-216.
[7] Li yuexin, Hu jie, Qin li, Gu mengxia. Knowledge engineering base and applied case. Science Press. Beijing. 2006; 10.
[8] Thomas R Gruber. Toward Principles for the Design of Semantic network Used for Knowledge Sharing, Revision. 2005.
[9] ABM Nasiruzzaman, Complex Network Framework Based Comparative Study of Power Grid Centrality Measures. *IAES International Journal of Electrical and Computer Engineering.* 2013; 3(4): 124-138
[10] Li man, Wang yan, Zhao yiyu. Research of large-scale storage model Ontology based on relational database. *Journal of Huazhong University of Science and Technology*. 2005; 12: 217-220.
[11] Dimitris N Chorafas. Knowledge Engineering. New York: Van Nostrand Reinhold. 1990.
[12] Jin cong, Dai shang-pin. Artificial Intelligence tutorial. Beijing: Tsinghua University Press. 2007; 7.
[13] Xiong dong-min. Research of Chinese word segmentation and recognition of Chinese names. 2006; 3.
[14] Tang su-qin, Li-bo. Intelligent question answering system based on sentence template. *Journal of Guangxi Normal University*. 2007; 6.
[15] Liu jun. Knowledge inquiry system based on short message. 2006; 4.
[16] Yan wei-min, Chen wen-bo. Data structures and application algorithms tutorial. Beijing: Tsinghua University Press. 2008; 4.
[17] Zhu qin-ji. Research of implicit knowledge query in database query system. *Computer Engineering and Application*. 2002; 19.
[18] Dmitri L Chora. Through powerful interactive Web-enabled knowledge automation expert systems. 2010; 12: 234-258.