An Ontology-based Knowledge Matching Algorithm for the Strategic Study

Zhihao Shang, Lili Rong
Institute of Systems Engineering, Dalian University of Technology
Dalian, China, 116024
Email: llrong@dlut.edu.cn

Abstract

This paper is concerned with the question of how to provide knowledge support for the strategic study. An ontology-based matching algorithm is presented. The algorithm consists of four steps. First, we introduce how to build the knowledge tree. The knowledge tree is the explicit specification of the documents collected. Every document can be found in the nodes of the knowledge tree. The knowledge tree starts with extracting the framework of the books with respect to a subject and will be completed by hanging the documents on the nodes of the framework. Second, we introduce how to build the question tree. The question tree gives the explicit specification of one question by analyzing it. Third, the relations of the nodes between the knowledge tree and the question tree are discussed. Finally, different matching algorithms are proposed according to the relevant relations. The documents hung on the matched nodes of the knowledge tree can support the researchers on studying the strategic question.

Keywords: ontology, strategic study, knowledge tree, question tree, knowledge matching algorithm

1. INTRODUCTION

There are two methods of studying regional strategy. One is quantitative study, and the other is qualitative study. Historical and current statuses are analyzed and compared in qualitative study, and data is analyzed in quantitative study. Both quantitative study and qualitative study need the support of documents. Recently, there are many researches about how to work out the strategic plan, such as question leading method, target demonstrating method, main factors of comprehensive competitiveness analyzing method, multi-stage researching method, implementing strategy studying method and evaluating strategic environment method[1][2]. However, there are few researches about how to provide documents for strategic study. Recently, with the fast development of information technology, the documents are usually hunted out from web with some search engines. For one subject, hundreds of documents may be found out. A subject usually consists of many questions, so only tens of the documents may be related to a question of the subject. The researchers have to choose the documents corresponding to a question one by one from all documents. It takes too much time. So it will be very helpful to extracting the documents needed for the researchers. If we regard the documents correlated with the question as the demander of knowledge and the documents correlated with the subject as the supplier of knowledge, matching the demander with the supplier can be accomplished by matching the knowledge tree with the question tree.

Now ontology-based method is one of the good methods of solving the question of supply-demand matching. By building the domain ontology, the concepts in the domain can be described well. According to different domain, different ontologies are built to provide knowledge explicitly[3]. Recently, the ontology-based method has been used widely. Zbigniew[4] presented ontology-based distributed autonomous knowledge systems to handle semantics inconsistencies between the same attributes used at different sites. Storey[5] presented an ontology that can be used as a surrogate for the meaning of words in a database design system to simulate the contributions that a designer would make based on his or her general knowledge. Kiryakov Atanas[6] built a semantic web with an ontology. However, the ontology-based method is seldom used to match the demander with the supplier for the strategic study.

So an ontology-based knowledge matching algorithm is presented for the strategic study in this paper. The algorithm consists of four steps. The first step is building the knowledge tree. This tree is the supplier of the knowledge. Based on the ontology, the documents gotten are specified with a tree. The second step is building the question tree. This tree is the demander of the knowledge. The researchers decompose a question into some subquestions and assemble the question and subquestions with a tree. The third step is analyzing the relations of the nodes between the question tree and the knowledge tree. And the last step is matching the knowledge tree with the question tree. The various relations of the nodes between the knowledge tree and the question tree are discussed in Section 3 below. Different matching methods are presented according to different relations. At last, some subtrees will be extracted from the knowledge tree. The documents hung on these subtrees can provide knowledge support for the researchers.

2. DESCRIBE THE QUESTION AND THE KEY TECHNOLOGY

2.1. Describe the question

The knowledge set, the question tree and the target set are discussed in this section. The relations of them are illustrated in Fig. 1.

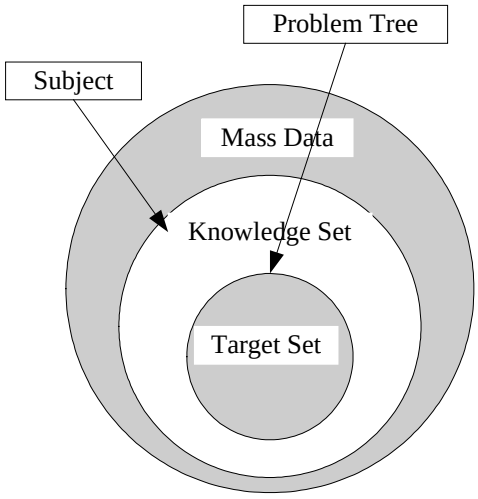

Fig. 1. The relations of three concepts

The knowledge set aggregates all of the documents with respect to the subject. Some documents are gathered from mass data with the search engines. These documents are assembled and classified that a set of knowledge is formed.

The question tree gives an explicit account of a question of the subject. It consists of a question and its subquestions, and it can describe a question concretely.

The target set aggregates all of the documents with respect to the question. With some matching algorithm, the question tree is matched with the knowledge set. The result of matching contains the needed documents which form the target set.

2.2. Ontologies

The word ontology was taken from Philosophy, where it means a systematic explanation of being. It emphasizes the nonobjective essence of being. In the last decade, the word ontology became a relevant word for Knowledge Engineering community. There are many definitions of ontology. One of the first definitions was given by Neches and colleagues[7] as follows: “ an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combing terms and relations to define extensions to the vocabulary”. In 1997, Borst[8] defined an ontology as follows: “Ontologies are defined as a formal specification of a shared conceptualization”. Gruber defined an ontology as “an explicit specification of a conceptualization”. This definition became the most quoted in literature and by the ontology community.

3 DESCRIBE THE SUPPLIER AND DEMANDER OF KNOWLEDGE

3.1. Describe the supplier of knowledge

Describing the supplier of knowledge means building a knowledge tree. A knowledge tree expresses the knowledge set with a tree. Fig. 2 shows a model of a knowledge tree. Building a knowledge tree is based on the concept of the ontology. Like an ontology, a knowledge tree is also built to give an explicit account of

each document.

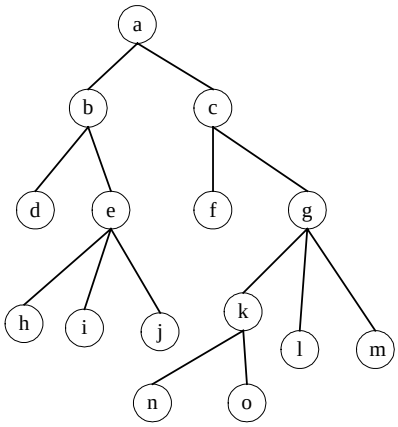

Fig. 2. Model of a knowledge tree

Describing the supplier of knowledge starts with building a framework of knowledge. Building the framework of knowledge consists of three steps. The first step is to choose some books by the experts. These books must be important and representative in the subject. The second step is to extract the frameworks of these books. The third step is to combine these frameworks to form a big framework. We can choose the most comprehensive framework from these frameworks. Then we perfect this framework by referring to other frameworks. The big framework can be regarded as the rudiment of the knowledge tree.

Building the framework of knowledge should also be based on ontology. An ontology is considered as a theory which gives an explicit, partial account of a conceptualization. A knowledge tree is also built to give an explicit account of each document. So a knowledge tree is similar to an ontology.

After building the rudiment of the knowledge tree, the documents should be hung on the nodes of the knowledge tree. The content of a node describes one aspect of the subject. Hanging the documents on the nodes means matching the documents with nodes. In order to match a document with a node correctly, not only the content of the node but also the content of parents of the node should be considered. One document may correlate to more than one aspects of the subject, so one document can be hung on more than one nodes. If one document can be hung on a node and it can also be hung on a child node of the node, this document should be hung on the latter one.

3.2. Describe the demander of knowledge

Describing the demander means building a question tree. Each problem has its own structure. And a problem can be expressed with a group of phases and key words. According to the structure of the problem, these phases and key words are assembled to make up of a question tree. Fig. 3 shows a model of a question tree.

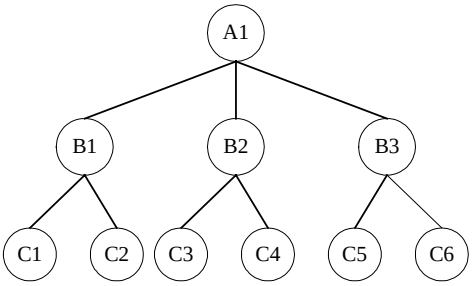

Fig. 3. Model of a question tree

3.3. Build the relations of the nodes between the question tree and the knowledge tree.

A question tree is built to denote a question. There are many different relations of the nodes between the question tree and the knowledge tree. Five basic relations are extracted as follows below.

Relation 1, Node Parent and Node I are matched, and Node A and a child node of Node I are matched. This relation is illustrated in Fig 4. In this figure, node *C6* represents Node A, node *B3* represents Node Parent, and node *e* represents Node I.

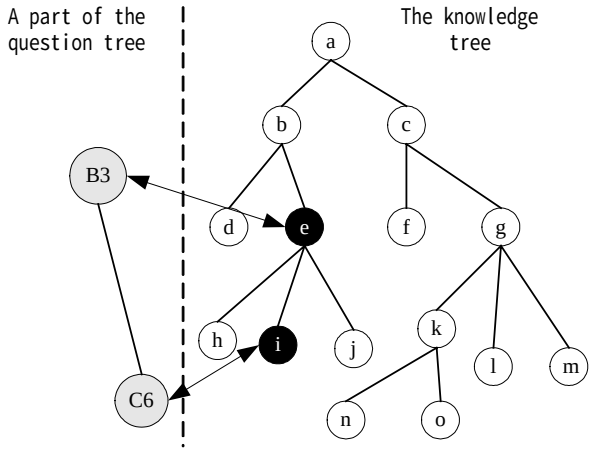

Fig. 4. Model of Relation 1

Relation 2, Node Parent and Node I are matched, and Node A is matched with a node which is neither a parent node of Node I nor a child node of Node I. This relation is illustrated in Fig 5. In this Figure, node *C4* represents Node A, node *B2* represents Node Parent, and node *k* represents Node I.

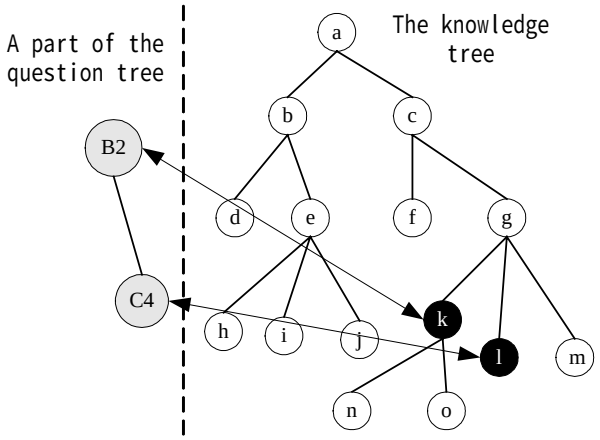

Fig. 5. Model of Relation 2

Relation3, Node A is matched with a node of the knowledge tree, but no nodes are matched with Node Parent. This relation is illustrated in Fig 6. In this Figure, node *C2* represents Node A, and node *B1* represents Node Parent.

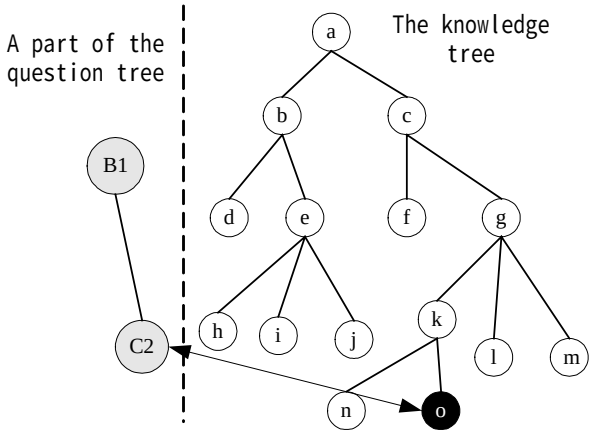

Fig. 6. Model of Relation 3

Relation4, Node Parent and Node I are matched, and Node A is matched with a parent node of Node I. This relation is illustrated in Fig 7. In this Figure, node *C5* represents Node A, node *B3* represents Node Parent, and node *e* represents Node I.

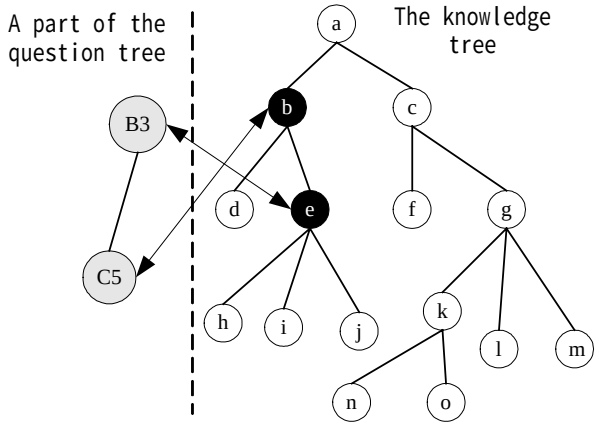

Fig. 7. Model of Relation 4

Relation5, Node Parent is matched with a node of the knowledge tree, but no nodes are matched with Node A. This relation is illustrated in Fig 8. In this Figure, node *C3* represents Node A, node *B2* represents Node Parent, and node *k* represents Node I.

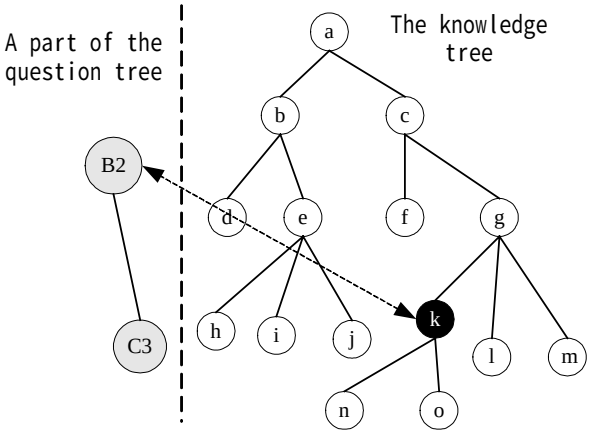

Fig. 8. Model of Relation 5

Where, Node A denotes a leaf node of the question tree, Node Parent denotes a parent node of Node A, and Node I denotes the node which is matched with Node Parent in the knowledge tree.

4. MATCH THE KNOWLEDGE TREE WITH THE QUESTION TREE

Based on the above five relations, six matching algorithms are presented. Five of them correspond to five basic relations. Another matching algorithm corresponds to the combined relation which is composed of more than one basic relations. Fig. 9 shows a

combined relation. This relation consists of Relation 1 and Relation2. A matching process is shown in Fig. 10. The six algorithms are contained in the process.

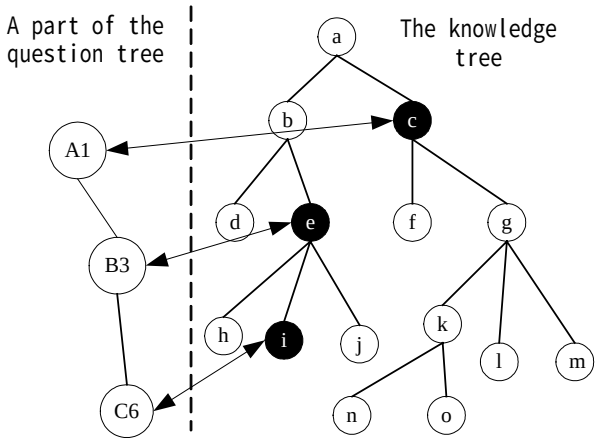

Fig. 9. Model of a Combined relation

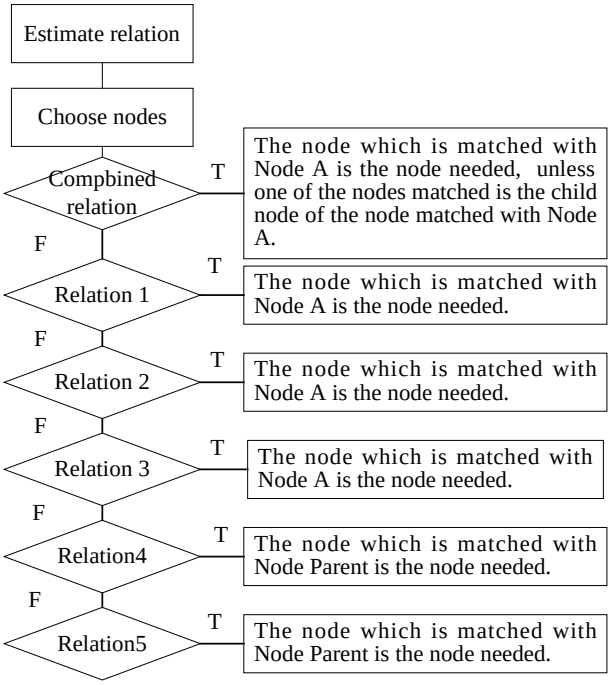

Fig. 10. Process of matching the question tree with the knowledge tree

In Fig. 10, *Choose nodes* means choosing a leaf node of the question tree first and choosing all of parent nodes of the leaf node then.

With the algorithms, the results to the relations shown in the Figures above (Fig. 4-Fig. 9) are presented below. In Fig. 4, node *i*(Node A) is the needed. In Fig. 5, node *l*(Node A) is the needed. In Fig. 6, node *o*(Node A) is the needed. In Fig. 7, node *e*(Node Parent) is the needed. In Fig. 8, node *k*(Node Parent) is the needed. In Fig. 9, node *i* is the needed.

The subtrees whose root nodes are matched with the nodes of the question tree are assembled into a new tree in order. The documents hung on the nodes of this new tree can provide the researchers knowledge support to study strategic questions.

5. AN INSTANCE

With the strategic study of sustainable development of the old industrial base of Liaoning province as an example, how to match the supplier with the demander is presented. A knowledge portal for sustainable development of the old industrial base of Liaoning province should be built first. The knowledge portal is an intelligent web station which serves some special people and the limited domain. It sets up the relationship between users and knowledge. It helps users find relevant knowledge or knowledge media. The users can also communicate with each other in the portal. There are six parts in this knowledge portal. They are *Provincial condition*, *News*, *Actuality of development*, *Focal question*, *Statistical data*, *Law*, *Relevant literature*, *Aim*. The documents are collected according to these six items. The numbers of documents of all items are listed in Table 1. Four books are also referred to.

Table 1. Numbers of documents of all items

Type	No.
Provincial condition	27
News	15
Actuality of development	36
Focal question	14
Statistical data	62
Law	16
Relevant literature	145
Aim	13

The book usually contains the most comprehensive content correlated to one subject. One book, Embrace Spring of the Market[9], is selected by experts. The structure of this book is extracted and the rudiment of

the framework of knowledge is gotten. By inferring to other books and articles, this framework is perfected. Three are two hundreds and twenty-two documents hung on this tree. Fig. 11 shows a part of the knowledge tree.

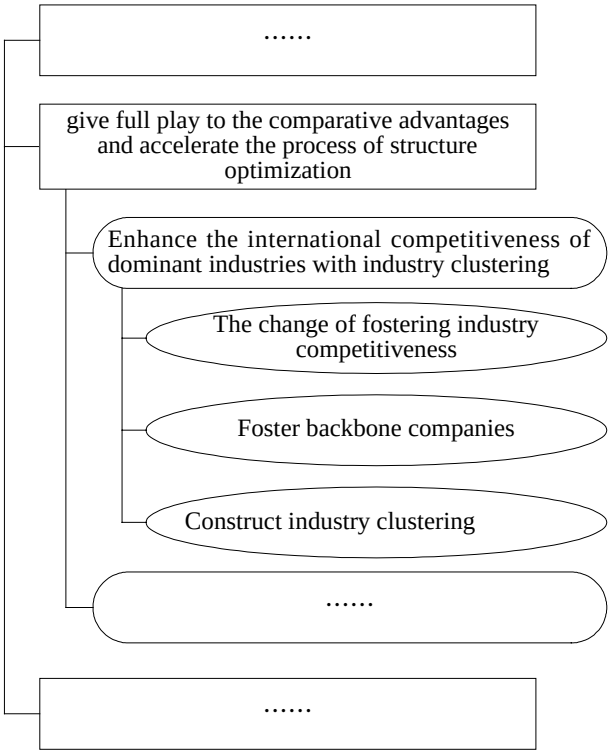

Fig. 11. A part of the knowledge tree

One question, *environment-adapting ability*, should be discussed in the strategic study. This question can be decomposed into four subquestions: *globalization*, *knowledge-intensive*, *binning* and *informationazation*. So a two-layer tree is gotten (see Fig. 12). The relation of the nodes between the question tree and the knowledge tree belongs to Relation 3 as shown in Section3.

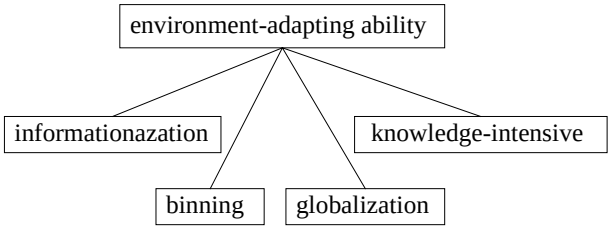

Fig. 12. The question tree

After matching the question tree with the knowledge tree, we can get the result as shown below: There are two subtrees matched with *globalization*: the subtree whose root node is "the well-to-do society and vitalizing old industrial case of the international economic trade of informationazation, globalization and multipolarization" and the subtree whose root node is "implementing international resource strategy is the impersonal request of economic globalization". There are three subtrees matched with *binning*: the subtree whose root node is "the well-to-do society and vitalizing old industrial case of the international economic trade of informationazation, globalization and multipolarization", the subtree whose root node is "northeast asian regional economic cooperation is the inevitable choice of vitalizing northeast old industrial case" and the subtree whose root node is "construct industry clustering". There is only one subtree matched with *informationazation*, and its node root is "informationazation is the only way to equipment manufacturing industry, high technology industry and raw material industry". None of the subtrees are matched with *knowledge-intensive*.

Twenty-seven documents are chosen from two hundreds and twenty-two documents, so that the time of finding the knowledge needed is saved.

In order to prove this arithmetic efficient, four experts are invited to evaluate on the basis of two questions. One question is whether the documents can support the question (Index 1), and the other question is whether the documents are enough (Index 2). To one question, an expert gives a mark. The highest mark is 5, and the lowest mark is 1. Table 2 shows all the marks.

Table 2. Result of evaluation

Expert / Index	Expert1	Expert 2	Expert3	Expert4
Index1	5	5	4	5
Index2	3	4	3	3

The result shows that the documents chosen can support the question, but the documents are not much enough. It is because the documents collected are limited at the beginning. The documents vary every day that you cannot get all the related documents.

6. Conclusion

In this paper, we have presented an ontology-based

matching algorithm for the special request of the strategic study. By building the knowledge tree and the question tree and matching them, we can provide the researchers the knowledge with respect to a certain question. This algorithm works effectively in the strategic study of sustainable development of the old industrial base of Liaoning province. In order to make this algorithm more efficient, how to complete hanging the documents on the nodes of the knowledge tree and matching the two trees automatically by the computer will be studied in future.

REFERENCES

[1] Heracleous, Loizos, Strategic thinking or strategic planning, *Long Rang Planning* 31(3) (1998) 481-486

[2] Linda Candy, Representations of Strategic Knowledge in Design, *Knowledge-Based Systems* 11(7-8) (1998) 379-390

[3] Sugumaran Vijayan, Storey Veda C., Ontologies for conceptual modeling: their creation, use, and management, *Data & Knowledge Engineering* 42(3) (2002) 251-271

[4] Ra&sacute, Z.W.Zbigniew W., Dardzińska , Agnieszka, Ontology-based distributed autonomous knowledge systems, *Information Systems* 29(1) (2000) 47-58

[5] Storey Veda C., Dey Debabrata, Ullrich Harald, Sundaresan Shankar, An ontology-based expert system for database design, *Data & Knowledge Engineering* 28(1) (1998) 31-46

[6] Kiryakov Atanas, Popov Borislav, Terziev Ivan, Manov Dimitar, Ognyanoff Damyan, Semantic annotation, indexing, and retrieval, *Web Semantics: Science, Services and Agents on the World Wide Web* 2(1) (2004) 49-79

[7] Robert Neches, Richard Fikes, Tim Finin, Thomas Gruber, Ramesh Patil, Ted Senator, William R. Swartout, Enabling technology for knowledge sharing, *AI Magazine* 12(3) (1991) 36-56

[8] Corcho Oscar, Fernández-López Mariano, Gómez-Pérez Asunción, Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & Knowledge Engineering* 46(2003) 41-64

[9] Zhao Xinliang, Embrace Spring of the Market, Liaoning People Press 2004 (in Chinese)