WUJNS
Wuhan University Journal of Natural Sciences

Vol. 11 No. 1 2006 078-082

Article ID:1007-1202(2006)01-0078-05

Construction and Resource Locating of Semantic P2P Grid Based on Description Logics

□ SUN Xiao-lin, LU Zheng-ding†, LI Yu-hua, WEN Kun-mei, LI Rui-xuan
College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

Abstract: This paper proposes an algorithm applied in semantic P2P network based on the description logics with the purpose for realizing the concepts distribution of resources, which makes the resources semantic locating easy. With the idea of the consistent hashing in the Chord, our algorithm stores the addresses and resources with the values of the same type to select instance. In addition, each peer has its own ontology, which will be completed by the knowledge distributed over the network during the exchange of CHGs (classification hierarchy graphs). The hierarchy classification of concepts allows to find matching resource by querying to the upper level concept because the all concepts described in the CHG have the same root

Key words: semantic P2P; description logics; consistent hashing; ontology; CHG (classification hierarchy graphs)
CLC number: TP 302.2

Received date: 2005-04-26
Foundation item: Supported by the National Natural Science Foundation of China (60403027)

Biography: SUN Xiao-lin(1980-), male, Ph. D. candidate, research direction: data mining, description logics, semantic Web. E-mail: arthursum2005@hotmail.com
† To whom correspondence should be addressed. E-mail: zdlu4409@public.wh.hb.cn

0 Introduction

Nowadays, the resource is used in a very broad manner. Various kinds of resources that are available have got an explosive growth. At the same time, with the appearance of ontology and semantic web, the request to inquiring details of resources has become more and more complicated, which means the analysis to semantic elements is needed sometimes. Thus, the resource locating and matching in P2P grid are getting more and more difficult.

the Semantic Grid initiative[1] aims at incorporating semantic technologies like ontologies and description logics into Grid middleware. Several approaches have been made to use ontologies in the matchmaking problem[2-4], but they use ontologies only locally and assume that the provider and requestor use the same background knowledge. This means that the nodes of resources must share the same knowledge base (TBox, ABox), but at the same time because the knowledge base is expanded and increased dynamically, it is more difficult to put the intact unified knowledge base in all the resource nodes to guarantee up-to-dateness and consistency of the ontology. So some researches based on ontology for P2P network resource locating and matching propose that distributed peer should use the distributed knowledge base[5]. This means all peers can going on reasoning and route searching alone.

Within this paper we give an overview about related methods with a brief introduction of description logics (DLs), distributed hash tables and Chord Algorithm in Section 1, these three technologies are the foundation of our algorithm, which is followed by the design of foundation and format of the DL system held in every peer in Section 2. In Section 3 we describe the core algorithm for retrieving peers with locating resources. In the remainder of the paper we present the conclusion and giving an outlook to our upcoming next steps.

1 Methods Related

There are many approaches of the research on the areas of reasoning systems and distributed peer-to-peer computing that we use as a foundation of our work. In the following, we introduce these methods, namely description logics, distributed hash tables and Chord Algorithm.

1.1 Description Logics

Description logics is the most recent name for a family of knowledge representation (KR) formalisms that represent the knowledge of an application domain (the "world") by first defining the relevant concepts of the domain (its terminology), and then using these concepts to specify properties of objects and individuals occurring in the domain (the world description).

Classification of concepts determines sub-concept (super-concept) relationships (called subsumption relationships in DLs) between the concepts of a given terminology, and thus allows one to structure the terminology in the form of a subsumption hierarchy.

A KR system based on description logics provides facilities to set up knowledge bases, to reason about their content, and to manipulate them. Figure 1 sketches the architecture of such a system. A knowledge base (KB) comprises two components, the TBox and the ABox. The TBox introduces the terminology, i.e., the vocabulary of an application domain, while the ABox contains assertions about named individuals in terms of this vocabulary[6].

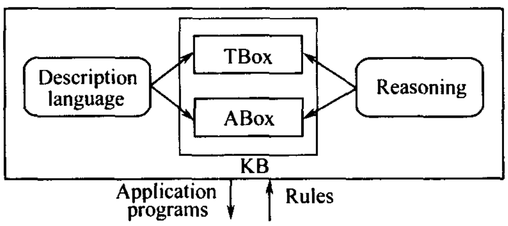

Fig. 1 Architecture of a knowledge representation system based on description logics.

1.2 Distributed Hash Tables

Distributed hash table (DHT) means each peer in the network has its own hash table. It can be easy to extend for the P2P grid applying this technology. Second generation P2P networks are based on distributed hash table algorithms. These algorithms allow locating a peer in a P2P network, which holds some information for a specific key. There are different DHT algorithms, we use currently the constant hashing algorithm Chord[7], which allows to find the target node in $O(\log(n))$, where "n" means the size of the network.

Chord is a kind of DHT algorithm to distribute nodes and resources in P2P with different identifiers. It applies a varietal method of consistent hashing value computing. There are several advantages in consistent hashing algorithm. Consistent hashing can distribute every node and resource keyword with unique identifier which produced by hash algorithm. The identifier of node can be count by hashing the IP address of the node. By contrast, we can get the identifier of keyword by hashing itself.

2 Description Logics System in P2P Grid

This algorithm is based on the DL system hold in different nodes that construct the distributed P2P network. The CHG is the important part of the distributed DL systems to keep their structures uniform.

2.1 Classification Hierarchy Graph

Every peer runs a local description logics system and uses it to store its own resource catalogue. It describes the existence of certain resources by making A-Box assertions. Every resource is described as an individual, which is asserted to be a member of one or more concepts. These concepts describe all aspects of the resource. Because the P2P grid is an open system, it is difficult to force every peer to use the same vocabulary or TBox.

We use classification hierarchy graph (CHG) of concepts to make different peers work together in the case of distributing TBoxes. As we mentioned in Section 1.1, DL provides subsumption relationships between concepts, sub-concept and super-concept. From this subsumption relation the so-called classification graph is calculated, which is built by taking the concepts as vertices and the subsumption relation as directed edges. There is a universal concept in each DL system to subsume all concepts declared in DL. We refer to concept "Resource" as the root node of the classification tree and this acyclic concept classification hierarchy graph as CHG as it shown in Fig. 2.

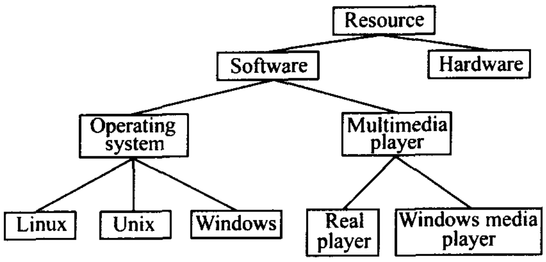

Fig. 2 Classification hierarchy graph

When a node joins in the network, it will send its local CHG to an other node which is the closest member of the P2P network. Since the different node in the same P2P system has the same CHG, the node put into newly can receive a unified CHG from the closest P2P network member. In the exchange process of CHGs, if the CHG in new node conflicts to the local CHG, the node of P2P network will check the soundness of CHG according to local DL system. If local CHG is errorless, the correct CHG should be sent back to the node applying to join, otherwise it will send error reporting and the information for revising CHG to all nodes of P2P network.

In this way, we can guarantee the consistency of the concept structure in P2P network , which will contribute to the search and locating for resources. What we should pay attention to here is, the same concept may be distributed to different nodes, thus, though the hierarchy graph of the concept is an acyclic arborescent structure, relations between the nodes will be much complicated.

2.2 Distribute Knowledge Base

The knowledge base in distributed DL system has two components as TBox and ABox. Our goal is to provide a system which allows distributed instance check in a peer-to-peer network where both the ABox and the TBox are distributed over the network.

In our model, Each peer has its own knowledge base, as it is shown in Section 1.1, TBox and ABox all have their respective forms. We define one form of terminology named subsumption relationship to show the relation of nodes as the main content of distributed TBox. Its form is as follows:

$$\text{Concept } X \subseteq \text{Super - concept } Y$$
$$\text{Sub - concept } A_1 \subseteq \text{Concept } X$$
$$\text{Sub - concept } A_2 \subseteq \text{Concept } X$$
$$\cdots$$
$$\text{Sub - concept } A_n \subseteq \text{Concept } X$$

X represents the concept of the node, and Y represents the super-concept of the node. We prefer to use "super-concept" and "sub-concept" to show the concept that has direct subsumption relationship with the target node, while the indirect deducible subsumption relationship can be expressed as "super-super-concept" or "sub-sub-concept". For example, if Concept A is the super-concept of concept B, at the same time, Concept B is the super-concept of concept C. Then we call Concept A is the super-super-concept of concept C. As we have mentioned above, we could find one super-concept at most in CHG. Therefore, the name of super-concept and the list of sub-concept can be found in TBox of the concept node.

In ABox, the list of resources $(k_{m+1}, k_{m+2} \cdots k_{m+n})$ that local node hold will be stored. Besides these, a node should save the hash values of IP addresses for its super-concept $(k_1, k_2 \cdots k_m)$ and sub-concepts (k_{m+n+1}) nodes. We indicate these assertions as following format:

$$X(k_1), X(k_2), \cdots, X(k_m)$$
$$A_1(k_{m+1}), A_2(k_{m+2}), \cdots, A_n(k_{m+n}), Y(k_{m+n+1})$$

In addition, with the idea of DL, we put a certain basic rule for each knowledge base. For example, the "k" in resource instance assertion $X(k)$ is not allowed to appear in other resource instance assertion of this node, or the super-concept must be defined and no more than once in TBox, etc.

2.3 The Construction of Semantic P2P System

In this section, we will represent the course of construction based on P2P platform of distributed DL system through the examples of peer joining and leaving.

2.3.1 Peer joining the P2P network

When a node wants to be a peer in the network, the first thing it should do is to determine its concept name by the underlying local DL system. Figure 3 shows the details of the TBox and ABox in a node named "Windows" (node 1). The value 22 is counted as the hash value of the resource keywords by DHT algorithm just like Chord. The value 17 is the consistent hashing value for the IP address of this node describing the concept "Windows". Then the node with the concept name "Windows" sends its CHG to the closest node it can reach in

Windows(17)
TBox: Windows ⊆ OperatingSystem
ABox: Windows(22)

Fig. 3 Node representing "Windows"

the P2P grid, and communicates with it to decide the correct CHG. As the two results we have discussed in Section 2.1, its CHG is contented as same as the target P2P network.

To find its location in the network, the node "Windows" must send two kinds of message to nodes in the P2P network as it is shown in Fig. 4 and 5. There is some information taken by the first kind of massage, which is the name of the super-concept. The peers share the same super-concept of the node "Windows" named "Operating System" (node 2) will receive this message and send back answers about the concept "Windows" that already exists in the network if the "Windows" is not a new concept here. When node 1 finds the same concept here it will send the second kind of message to the closest one. The message including the local resource list of "Windows" would be sent to "Windows" node that already exists (node 3) to ask for the list of sub-concept

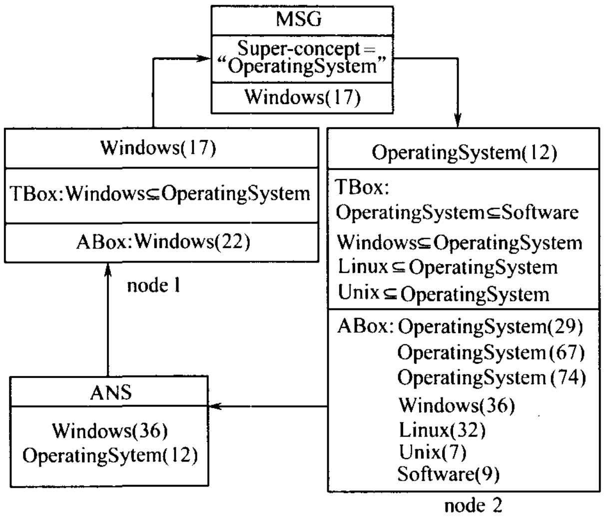

Fig. 4 Message sending step 1

of concept "Windows". Node 3 will check the resource instance list it receives, and then combined it with the local resource information to make an integrated list. At last, node 3 will send it back to node 1 which wants to join in with the list of sub-concepts to realize the node in same concept have the same knowledge bases in P2P network. If concept "Windows" is a new one that did not be defined in the network before, node 1 need to revise P2P CHG in the network instead of seeking the existence of the node with same concept name for the list of resources and sub-concepts after sending first kind of message.

Figure 6 shows the details of 3 nodes after node 1 joining in. We can find node 1 and node 3 have the same

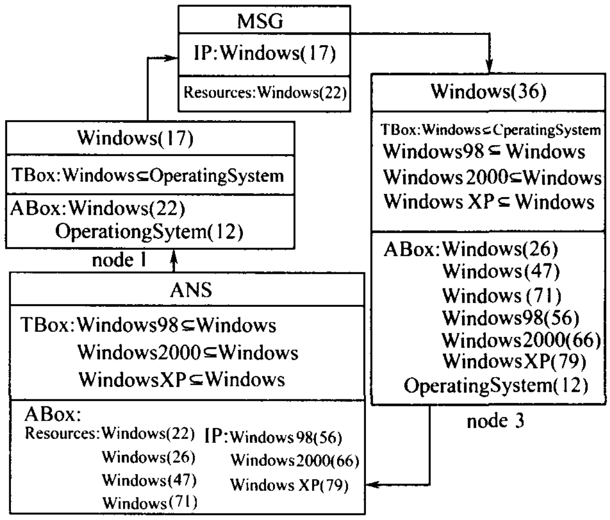

Fig. 5 Message sending step2

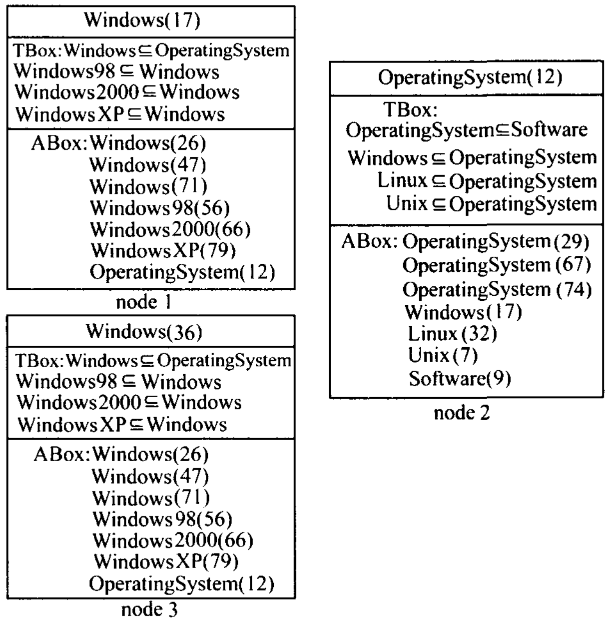

Fig. 6 Details of 3 nodes after messages flow

TBox and ABox, and the assertion about concept "Windows" has become Windows (17) in the ABox of node 2.

2.3.2 Peer leaving the P2P network

We consider the situation that peer leaves normally here only. The unexpected situation in leaving from the network is beyond our research. The course peer leaves and the course it joins are very similar. After sending messages to its super-concept node at first, the leaving peer will then inform the super-concept whether it is the last node of this concept. The super-concept makes different responses according to the different cases. If there are nodes of this concept exist in the network yet, super-concept node need to send the revising information about resource list to the rest all sub-concept nodes and amend its own ABox. Otherwise the super-concept only modify one's own ABox to remove the information about the node left. Anyway, it is not necessary to revise the CHG and TBox in any node, because the generally terminology is always correct and the concept will exists forever, even there is no instance to explain it.

3 Locating the Target Node

After introducing the structure of the ontology-driven P2P grid, we will explain that how the peers locate the resource they need in this network by semantic querying.

We continue to use the example mentioned above to demonstrate the cause of node locating. If the node 1 who representing "Windows" wants to find the resource of a kind of multimedia player software named "RealPlayer", the local CHG should be queried at first. Because the CHG is a tree with the root named "Resource", it is capable of finding the concept that contains the target concept "RealPlayer" and "Windows" both, and then chose the one that is closest to the target in the CHG. In our example, node 1 will find concept "Software". Node 1 can locate the one of the peers that describes the concept "Software" by inquire upwards ceaselessly. The node located by the node 1 will get a message from it, which has a stack of concepts containing the target concept. It is "Software", "multimedia player" and itself "Real Player". The "Software" node pops the first element named "Software" and sent this stack to all its sub-concepts and one of them will be satisfied with the second element named "multimedia player". Then repeat this step till finding the target node. The one of "Real Player" nodes will answer the query to node 1 according to the hash value of its IP stored in the message. Figure 7 shows that how the node 1 find the resource "Real Player" that he need.

4 Conclusion

In this paper a novel approach is presented for distributed resource locating in the Grid. Our system uses distributed ontologies based on description logics to describe the resources. It distributes the information over a peer-to-peer network based on distributed hash tables.

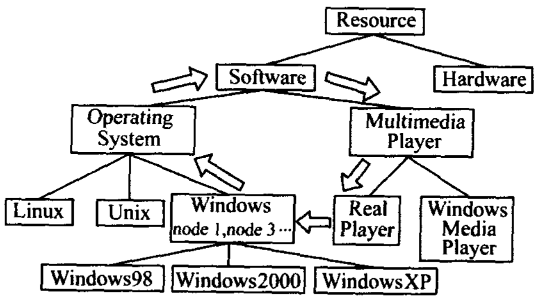

Fig. 7 Process of querying

We outline the core algorithm and depicted how it enables the system to unite the distributed TBox, and guarantee its soundness. This has been achieved by combining the knowledge of all peers within a distributed CHG. Our algorithm ensures the data stored in different peers become average, which makes each node realize being equal.

However, there is little discussion about reasoning algorithm in this paper, although the reasoning algorithm is the most important part of the DL system. That will be one of our future work. Besides this, Completeness of ontology-based system, Expressiveness of queries, Fault tolerance and accident checking will be the next emphasis in our research.

References

[1] Goble C, Roure D D. The Grid: An Application of the Semantic Web. *SIGMOD Rec*, 2002, **31**(4):65-70.

[2] Brooke J, Fellows D, Garwood K, *et al*. Semantic Matching of Grid Resource Descriptions. *Proceedings of the 2nd European Across-Grids Conference*, Nicosia, Cyprus: University of Cyprus, January 2004. 72-81.

[3] Tangmunarunkit H, Decker S, Kesselman C. Ontology-Based Resource Matching in the Grid-the Grid Meets the SemanticWeb. *Proceedings of SemPGRID '03*, Budapest, Hungary. May 2003. 85-101.

[4] González-Castillo J, Trastour D, Bartolini C. Description Logics for Matchmaking of Services. *Proceedings of the KI-2001 Workshop on Applications of Description Logics*, Germany: University of Ulm, September 2001. 44-67.

[5] Felix H, Matthias H. Towards Ontology-Driven P2P Grid Resource Discovery. *Proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing*, 2004, **32**:76-83.

[6] Baader F, Calvanese D, McGuinness D, *et al*. *The Description Logic Handbook*. Britain: Cambridge University Press, 2003. 47-73.

[7] Stoica I, Morris R, Liben-Nowell D, *et al*. Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications. *IEEE Transactions on Networking*, 2003, **11**(1):17-32.

□