Research on a Mechanism for Assisting Service Retrieval

Bin Tang, Leqiu Qian, Ruzhi Xu,Yunjiao Xue
Department of Computer Science and Engineering, Fudan University, Shanghai, China
{tangbin, lqqian, yjxue, rzxu}@fudan.edu.cn

Hui Tang
Technical Support Section, Business Support Center, Guangdong Mobile Communication Co., Ltd., Guangzhou, China

Abstract

Through as an important part of service-oriented system architecture, service retrieval receives much attention from industry and academia and many methods have been proposed, there is still much to be expected of the recall, precision, efficiency and easiness of service retrieval. One important cause of this is that users cannot define their service retrieval requirements precisely and completely and this limits the function of ontology service models. Hence, an efficient service retrieval system should be able to support partial matching and help users to refine their retrieval requirements. The paper introduces a mechanism to assist users in giving high quality retrieval requirements. The goodness of the introduced mechanism to the recall, precision, efficiency and easiness of service retrieval is analyzed and its feasibility and effectivity is also verified through experiments.

1. Introduction

As an important part of service-oriented system architecture, service retrieval receives much attention from industry and academia. The performance of a service retrieval system is measured from the recall, precision, efficiency and easiness [9]. In the paper efficiency is broadly measured by the time from the users beginning to give a retrieval requirement to their selection of the services needed finally, which can be divided into three parts: the time users use to give a retrieval requirement, the time the system use to retrieve services based on the given retrieval requirement and the time the user use to select the services from the result retrieval set fit their need. Easiness is the measured by the efforts needed to retrieve the needed services.

According to the richness of the information contained in the service description model, web service description and retrieval model can be divided mainly into two categories:

- Syntactical level

In description, this category of models emphasizes the Syntax of the interface of web service and lack of constraints of behavior. In retrieval, them mostly based on key words. The representative systems based on this category of models are the UDDIs of IBM, Microsoft, SUN, which are simple to realize but the retrieval effect are not desirable [6][8].

- Semantic level

This kind of service models introduces ontology to describe services thoroughly and provide description of the function and behavior, which can be expected to give good retrieval result. The match of retrieval requirement and service advertisements can be based on logic reasoning and the result can be divided into five classes according to the matching degree: exact match, generalized match, specialized match, partial match and no match. There are also researches to calculate matching degree of retrieval requirement and service advertisements based on similarity counting, which calculate the matching degree of them as the weighted sum of the similarity degree of the property values in the retrieval requirement and service advertisement [1] [5]. The latter kind of methods to calculate similarity, which is adopted in the paper, has higher efficiency and recall and can deal with incomplete retrieval requirement and can let the user better know how the retrieved components fit his need. Representative studies include the augment UDDI Registry of Carnegie Mellon University [6], the Meteor-s project University of Georgia [5], the OWL-S ontology service standard of W3C [3], WSMO working group's WSMO ontology service standard [7] and so on.

Usually existing service retrieval methods have the assumption that users can well define their component queries. However, this assumption is not always realistic. Incomplete and imprecise retrieval requirements can heavily influence the quality of the retrieval result. The causes for user to give high quality retrieval requirements can come from three sources: service users usually does not know thoroughly of the service description model especially if the service model is described by an ontology model, which is a kind of most advanced service model expected to have good re-

Proceedings of The Sixth IEEE International Conference on Computer and Information Technology (CIT'06)
0-7695-2687-X/06 $20.00 © 2006 IEEE

trieval effect but a complex model too, this problem becomes more critical as the number of available services is dramatically increasing and so does the number of services with similar functions which cause service model became complex increasingly; Second, People may lack clear ideas about what they need while they begin searching for services; Third, service users may prefer to provide their initial retrieval requirements based on their necessary requirements, and hesitate to add the optional or unnecessary requirements into their retrieval requirements in order to avoid excluding possibly appropriate services. Because of the looseness of initial queries and the large number of available components, users usually still get numerous candidate components after providing their initial queries. In this situation, they need to refine their requirements step by step to narrow down the candidate components or exclude the unsuitable components. It is should introduce an assistance mechanism to help user to do this. The paper introduces such an assistance mechanism to assist users in giving high quality retrieval requirements. Below an ontology service model is described first, then the introduced assistance mechanism is outlined.

There are different kinds of ontology service models, in the paper we adopt the following model [1]:

WS=<CP, SP, Pre/Post, Is/Os, QoS>

Where CP are the common service properties, such as service name, application domain, the key words of the function of service, the category of service. SP are special service properties, which refer to concrete service's peculiar attributes and often domain dependant. Pre/Post are pre-condition and post-condition of services respectively. Is/Os attributes are the inputs and outputs respectively. Pre/Post and Is/Os are key properties to describe the functions of services. QoS is the quality properties of services, such as availability, security, performance, price, and reliability.

The paper introduces a mechanism to assist users in giving high quality retrieval requirements: Before retrieval, the possible implicit retrieval requirements is presented to users and added to the initial retrieval requirements under the users' confirmation. The association between implicit retrieval requirements and explicit retrieval requirements can be determined by domain expert or learn from retrieval history. After the first time retrieval, the possible ways to refine the property values of the retrieval requirements are obtained according to the difference of the corresponding property values of the retrieval requirement and the intermediate service retrieval result set and presented to the users (In service retrieval, service advertisements are described by the above service ontology model, and the paper adopt the query by example, i.e., the users' retrieval requirements are converted into a (potentially partial) description of the desired service using the above service ontology model). And the possible ways to complete the retrieval requirements is obtained according to the difference of the corresponding property values of the services in the intermediate service retrieval result set and presented to the users. According to the ways selected by users to refine property values of the retrieval requirements and to complete the retrieval requirements, the retrieval requirements is refined and service retrieval is carried out again in the intermediate service retrieval result set. If necessary, the service retrieval can be carried out for several times.

The paper is organized as following: the first part is the introduction, the second part of the paper introduced a mechanism for Assisting Service Retrieval and its goodness to the recall, precision, efficiency and easiness of service retrieval is analyzed, the third part demonstrates the feasibility and effectivity of the introduced service retrieval assistance mechanism through experiments, the last part gives the conclusion and feature work of the paper.

2. The mechanism for Assisting Service Retrieval

This part introduces a mechanism for Assisting Service Retrieval, 2.1 discusses the completion of users' initial retrieval requirement with implicit retrieval requirement based on association rules, 2.2 discusses refining of the property values of the of retrieval requirements, 2.3 discusses retrieval requirement completion based on the difference the services in intermediate retrieval result, 2.4 give the algorithm of the mechanism for Assisting Service Retrieval of the paper, 2.5 analyzes the introduced service retrieval assistance mechanism.

2.1 The completion of retrieval requirement with implicit retrieval requirements based on association rules

We can complete users' initial retrieval requirements based on the association between implicit retrieval requirements and explicit retrieval requirements. Possible implicit retrieval requirements is presented to users and added to the initial retrieval requirements under the confirmation. This kind of associations can be determined by the domain experts or through association mining algorithms. In 2.1.1 we will discuss the latter method. The paper does not introduces association mining due to limited space, interesting reader can refer to [2].

2.1.1 Adaptive service retrieval algorithm based on association rules. As in figure 3, the retrieval requirement of a user is sent to both the retrieval system and the learning system and the services selected finally from retrieval result by the users are also sent to the learning system. The learning system analyzes a amount of retrieval requirement and corresponding retrieval result periodically and obtains association rules of the form $A \Rightarrow B$, where A are the property values appeared in the user's initial retrieval requirements and B are the property values appeared in the descriptions of services in the retrieval result (In the association mining subsystem which learning the association between the explicit retrieval requirements and implicit retrieval requirements, each value of every CP or SP property is seen as an item, each value of Pre, Post, Is and Os are also seen as an item of the ontology item respectively. Each retrieval requirement and each service advertisements users selected finally is seen as a transaction), the probability of A and B appearing in the same one retrieval exceeds the given threshold minsupp, and the probability that the retrievals requirements contain A should also contain B exceeds the given threshold minconf. Then B is the requirement that the user probably needed.

The learning system delivers the obtained association rules to the retrieval system. In later retrieval, if A appeared in users' retrieval requirement (The right hand of each association rule is converted to a partial service description. If for each property existing in the corresponding service description of an association rule, the corresponding property in the retrieval requirements that are also converted into a (potentially partial) description of the desired service using the service ontology model as said in the first part have the same value, the association rule is determined to be the applicable), then the retrieval system adds the implicit B to the retrieval requirement under users' confirmation.

The following introduces the core algorithm of the association learning system.

Basic concepts: The user's retrieval requirement is denoted as Q, and there are m services in the retrieval result selected by the user, which are described by D_i respectively. The learning system maintains a database H of all past retrieval data, for every retrieval, the retrieval requirement Q and the D_i of every services selected by user are put into the database H.

The association rule of the retrieval requirement given by the user and the probable implicit requirement is of the form $A \Rightarrow B$, $A \cap B = \phi$. $A \Rightarrow B$ is satisfied in database H with support s and confidence c, the former is the percentage of the records containing $A \cup B$ in H, scilicet probability $P(A \cup B)$, the latter is the is the percentage of the records containing B in the set of records which contain A. The s should be larger than minsupp and the c larger than minconf.

The algorithm: Self-adaptive retrieval learning algorithm based on association mining adopts the Apriori to search the frequent attribute-value sets in the user's retrieval requirement and obtains information from the corresponding services selected by the user. The user's retrieval requirement Q and the description of each corresponding service selected by the user D are stored in the retrieval history database H and denoted as H_Q and H_D respectively. The method to obtain association rules of the algorithm is slightly different from wont in literature. The algorithm first utilizes the Apriori algorithm to select the set QL of frequent attribute-value sets contained in H_Q satisfying minsupp. For every set ql in QL, obtain the set DL of frequent attribute-value sets in the corresponding service selected by the user with frequency not less than the total number of the corresponding service selected multiply minconf. Select every set from DL and together with ql to form an association rule.

```
Algorithm: Self-adaptive Learning
Algorithm Based On Association Min-
ing
Input: retrieval history database
H, minimum support threshold min-
supp, minimum confidence threshold
minconf
Output: association rule sets R
(1)   Find frequent retrieval at-
      tribute-value sets QL in Q
      using Apriori
(2)   for ( i=1; QL•φ, i++ ){
(3)     fetch one frequent attrib-
        ute-value set ql in QL
        whose corresponding user
        feedback retrieval result
        set is D
(4)     find frequent attribute-
        value set DL whose support
        counting is larger than
        count( D)×minconf in D us-
        ing Apriori
(5)     Select every frequent at-
        tribute-value set dl from
        DL and together with ql to
        form a association rule
        ql⇒dl and put into R
(6)     remove ql from QL
(7)     },
```

Figure 1. **Adaptive service retrieval algorithm based on association mining**

The learning system sends the association rule set R

to the retrieval system which will add the property-value sets on the right side of applicable association rules to the retrieval requirement when carrying out a retrieval.

For the mobile phone message services, we obtain some association rules such as:

{application domain=enterprise}=>{operation log retrieval=true, message to a group=true, searching of ex messages=true }

{application domain= free commercial message service} $\Rightarrow$ {automatic divide of long message= false}

2.2 Refining of the property values of the of retrieval requirements

The property values of the retrieval requirements sometimes are also need to be refined. The system can help user to do this according to difference between corresponding property values of the retrieval requirement and the service advertisements in the intermediate result set. If a property value of a retrieval requirement is an ontology concept, it can have the following refinement ways: generalization, specialization, adding some attributes to a ontology concept, modifying the values of the attributes of a concept, or the attribute value should be the intersection of the current ontology concept and another ontology concept. For example, to retrieve online anti-virus services, there is special property of operating systems, if users assign windows to the property initially, windows can be refined to intersect with the concept server to windows server and can be specialized to windows server 2003 further. If a property value of a retrieval requirement is not an ontology concept but a word in WordNet, it can be refined based on the relations in WordNet to similar words of the corresponding property values of the services in the intermediate retrieval result. For the properties of key words of the function of service, Pre/Post and Is/Os, they can be seen as enumeration type, the possible refinement ways are to add a number to their value or refine a number of their value.

2.3 Retrieval requirements completion based on the difference of the descriptions of the services in intermediate retrieval result

If user is not satisfied with the intermediate retrieval result and the result is not empty, the retrieval system can return some problems, which are obtained by analyzing the difference of the corresponding property values of services in the intermediate result. And they can be mainly divided into two categories: one is about weather a property is needed or not because some services may have unique properties; another about what value a property take, to the latter kind of problems, the retrieval system provide the property values as candidate answers.

The user can give answers to the questions and the retrieval system refine the retrieval requirement based on the answers. The user can only answer some questions if he feels difficult and the system iteratively refines the result and there is an advantage: the settlement of some questions may make some other questions useless or the answers of other questions of simple. The problems can be arranged according to their importance. The importance of a question can be determined by experts according to the difficulty to answer the question and the information contain in the answer of the question. The importance of a problem can also be determined automatically. Let problem Pr is obtained based on property P, the values of P are P_1, P_2, …, P_m (P_m may equal empty means the corresponding service does not contain P), S_1, S_2, …,S_m are the sets of services in the intermediate result take P_1, P_2, …,P_m as the value of the attribute P respectively. We use the following information entropy to measure the importance of the problem based on attribute P ($p_i=S_i/S$):

$$I(S)=-p_1\log_2 p_1-p_2\log_2 p_2-\ldots-p_m\log_2 p_m$$

2.4 The algorithm of the mechanism for Assisting Service Retrieval

We have introduced three methods to help refine and complete retrieval requirement above and will give an algorithm of the mechanism for Assisting Service Retrieval based on the three methods as following:

Step 1 the user gives the initial retrieval requirement;

Step 2 the system complements the retrieval requirement based on the association rules obtained by the learning system after the user approval.

Step 3 the retrieval system obtain the result based on the retrieval requirement (if not the first time, the retrieval is carried out in the intermediate result), if the user is satisfied with the result goto step 7;

Step 4 return to the user the problems to refine the retrieval requirement and the possible ways of refinement of attribute values;

Step 5 refine the retrieval requirement based on the answers given to the problems and the ways selected to refine the values of the attributes of the retrieval requirements by the user.

Step 6 goto step 3;

Step 7 the end of the algorithm.

Figure 2. **Retrieval assistance algorithm based on feedback and refinement**

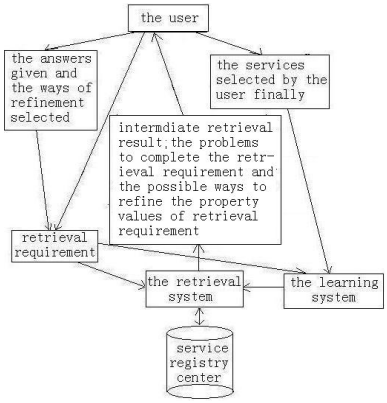

Figure 3. **The model of service retrieval system**

2.5 Analyzed of the introduced service retrieval assistance mechanism

The assistance and the goodness to the quality of retrieval result of the algorithm introduced in the paper are analyzed as following:

- Because of the refinement of the retrieval requirement based on association rules, the answers given to the problems and the ways selected to refine the property values of the retrieval requirements (possibly iteratively) by users with the assistance of the system, the precision of the retrieval result is improved.
- The easiness of the retrieval is embodied as following: relaxing the rigor requirement of the quality of the initial retrieval requirement given by the users; because the precision of the retrieval result is improved so it is easy for the user to select the services needed in the retrieval result.
- If define the service retrieval as including three phrases in a broad sense: the phrase the user gives an initial retrieval requirement and refine it with the help of the system, the phase the system retrieves service based on the retrieval requirement, the phrase the user selects the service needed, and because the participation of the user, classify the time after the first time of retrieval into the third phrase, because the first and third phase need much more time than the second phase and is reduced, so the retrieval efficiency in improved.

The recall of the retrieval is improved because the following two reasons: first, since the matching of retrieval requirement and services advertisements based on similarity computing and the services in the result set have a similar degree with the retrieval requirement greater than a threshold, the refinement of the property values in the retrieval requirement can improve the recall of the service retrieval; second, because the completion of service retrieval requirements, the services remain in the retrieval result better fit users' need.

3. Experiment

We have implemented a prototype service retrieval system called X-COM and verify the assistance mechanism introduced in the paper. We select one hundred and twenty services such as message services, data mining services and investigate the recall, precision and efficiency of JAXR Registry based on UDDI [10], the augment UDDI Registry that integrates DAML-S [6], the calculating the match of retrieval requirement and service advertisements based on the similarity of ontology and words Method I in [1] with the addition of Pre/Post and Method I in [1] with the introduced service retrieval assistance mechanism in the paper respectively. All systems run on windows 2003 and Eclipse SDK 3.1. (1) Register services described by WSDL according to NAICS standard to JAXR Registry, the retrieval is based on key words and according to NAICS standard; (2) Register services described by DAML-S to augment UDDI Registry, the retrieval is based on logic reasoning; (3) register services described by the ontology model given in the paper to X-COM, the retrieval is carried out based on the method I in [1] and result contains services the advertisement of which have a similar degree greater than the threshold, (4) register services described by the ontology model given in the paper to X-COM, the retrieval is carried out based on the method I in [1] and the retrieval requirement is refined based on the introduced mechanism in the paper and result contains services the advertisement of which have a similar degree greater than the threshold.

One hundred service retrievals are carried out for each system, the recall, precision and efficiency is calculated for every ten times. The average recall, precision and efficiency, which is defined as the time sum of the three phrase, are around 0.27, 0.21 and five minutes and ten seconds respectively for JAXR Registry, around 0.59, 0.53 and four minutes and fifty-seven seconds respectively for augment UDDI Registry, around 0.72, 0.66 and four minutes and thirty-five seconds respectively for method I in [1] and for method I in [1] with the introduced assistance mechanism the recall, precision and efficiency of service retrieval increase as the figure 4, 5, 6, respectively, where X-COM 1 represents method I in [1] with completion the retrieval requirements with implicit re-trieval requirements, X-COM 2 represents method I in [1] with refining of the property values of the of retrieval requirements and retrieval requirement completion based on

the difference the services in intermediate retrieval result and X-COM 3 represents method I in [1] with the whole introduced assistance mechanism.

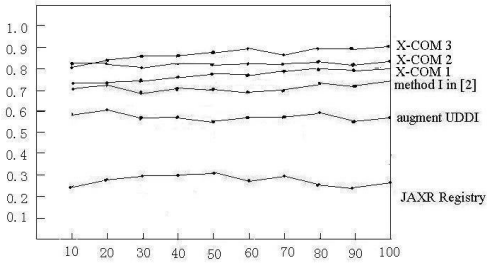

Figure 4. **The comparison of the recall**

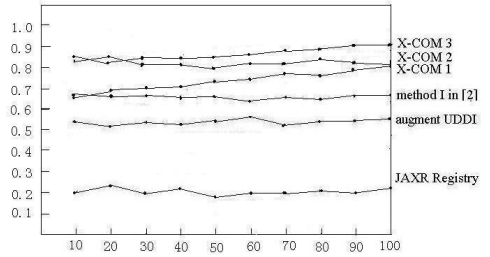

Figure 5. **The comparison of the precision**

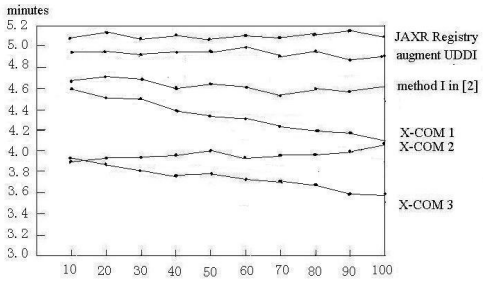

Figure 6. **The comparison of the efficiency**

4. Conclusion and feature work

The description and retrieval of services on syntactical level is easy to realize but the effect leave much to be expected. Ontology service model attempts to overcome this by utilizing ontology technology and there are many studies devote this. But users often cannot well define their component queries and heavily influence the quality of the retrieval result. It is urgent to have an assistance mechanism to assist users to give high quality retrieval requirement to make ontology service model into play better.

The mechanism introduced in the paper can assist users to refine and complete their retrieval requirement and improves the recall, precision, efficiency and easiness of a retrieval system, which is analyzed and demonstrated through experiments. The feature work is to optimize the efficiency of the algorithm and verify and improve the algorithm further.

Acknowledgements

This paper is supported by the National Natural Science Foundation of China under Grant No. 60473062, the National High Technology Development 863 Program of China under Grant No. 2004AA113030, Science Technology Committee of Shanghai under Grant No. 04DZ15022.

6. References

[1]Wu J, Wu CH, Li Y,Deng SG. Web Services discovery based on ontology and similarity of words. Chinese Journal of Computer, Vol.28, No.4, pp.595-602, 2005, 4.

[2]Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2000.

[3]W3C. OWL-S: Semantic Markup for Web Services. http://www.w3.org/Submission/OWL-S/, 2004.11

[4]Mark H. Burstein. Dynamic invocation of Semantic Web Services that use unfamiliar ontologies. Intelligent Systems, IEEE. Volume 19, Issue 4, pp.67-73, Jul-Aug 2004.

[5]Large Scale Distributed Information Systems Lab of the Georgia University, METEOR-S: Semantic Web Services and Processes, http://lsdis.cs.uga.edu/Projects/METEOR-S/ .

[6]Massimo Paolucci, Takahiro Kawamura, Terry R. Payne, Katia Sycara. Importing the Semantic Web in UDDI. In Proceedings of Web Services, E-business and Semantic Web Workshop, Toronto, Canada, pp.225~236, 2002.

[7]WSMO working group. D2v1.2. Web Service Modeling Ontology (WSMO). www.wsmo.org/TR/d2/v1.2/D2v1-2_20050414.pdf. 2005,4.

[8]Dogac A, Kabak Y, Laleci GB. Enriching ebXML registries with OWL ontologies for efficient service discovery. Proceedings of the 14th International Workshop on Research Issues on Data Engineering (RIDE'04), pp.69–76, 2004.

[9]Forbses Gibb , Colm McCartan , Ruairi O'Donnell , Niall Sweeney ,Ruben Leon. The integration of information retrieval techniques within a software reuse environment. Journal of Information Science, Vol.26, No.4, pp.211-226, 2000,4.

[10] http://www.sun.com/xml/jaxr.