Ontology-Driven Discovering Model for Geographical Information Services

ZHANG Lichao¹, WANG Keke¹, PAN Zhen¹,², WANG Qingshan¹, ZHENG Haiying¹
1. Institute of Surveying and Mapping, Information Engineering University, 66 Longhai Road, Zhengzhou 450052, China
2. First the Management Committee High & New Technology Industrial Development Zone of Zhengzhou, 6 Guohuai Road, Zhengzhou 450001, China

Abstract This paper proposes an ontology-driven discovering model for the geographical information services to improve their recall ratio and precision ratio. This model uses the geographical information service ontology. In this paper, first we study the multilevel matching arithmetic of geographical information services. This arithmetic is used for filtering and matching the services in the service register center according to the similarity between services selected and services requested from the definition of the function similarity and credit standing similarity. The matching arithmetic, geographical information service ontology and semantic description constitute the discovering model. Finally, we test and analyze the model from the recall ratio, precision ratio, responsivity and load balance. The result indicates that the ontology-driven discovering model is excellent in recall ratio and precision ratio, and can maintain the dynamic load balance of service copy.

Keywords ontology-driven; service discovering; function similarity; credit similarity; recall ratio; precision; load balance
CLC number P208; TP31

Introduction

Service matching is very important to discover the geographical information service. Semantics is used widely in service matching. Semantic Web service matching is realized through logic reasoning and glossary similarity. The glossary similarity arithmetic uses the glossary similarity of Information Retrieval to estimate the matching similarity of the function and performance description. Shi Zhongzhi matched and inferred the services through entering the description and reasoning of action into Information Retrieval. Mark Klein et al. put forward the high-precision service matching algorithm to discover services. They used the process model to catch the semantic, and then used the pattern matching arithmetic to discover the service to be selected. Katia Sycara put forward a very flexible service matching algorithm. He offered many matching strategies. The algorithm used the mechanism of semantic distance matching approximatively. However, the semantic distance depends on manual work. Therefore, there is a much questionable factor. All of these algorithms have the localization in practicability and dependability.

Ontology was originally a branch of philosophy that studies the essence of the existence of objective things. It is defined as the conceptual and specific canonical interpretation in this paper. The conjunction between symbols with material events is based on the common understanding to these relational conceptions in ontology. Ontology is the formalization structure
sustaining the knowledge sharing and reusing. It can be used to describe clearly the semantic of structured and half-structured information, and to achieve, maintain, access and save this information. Ontology provides the means to solve the isomerous expression of Web resources. The connotative domain model in ontology can be regarded as the semantic structure providing the uniform common expression\(^1\). The capability describing semantic and discovering resources provides the technology support and theoretical foundation for discovering geographical information services.

## 1 Ontology modeling

Ontology is commonly organized by taxonomy and expressed, described by formal language such as OIL, RDFS, OWL, etc. Ontology also can be described and saved by the notional pictures. Ontology depends on its description language. The formalization degree of the description language is higher and the automatic processing speed of the computer is quicker. OWL-S is the Web service ontology defined by the Web service language—OWL. It provides the key markup language structure for the Web service to describe the feature and function of the Web service clearly. OWL-S can describe the feature and function of the Web service through an accurate and readable form. A Web service is described by three features: presents, described by, and supports as the definition of the OWL-S. The bounds of the three features are the ServiceProfile, ServiceModel and Service-Grounding. The ServiceProfile is used to describe the ability of the Web service. It can not only be used to issue information by the service offers, but also describe the requirement of the service users\(^2\).

Geographical information service is used to complete basic geographical management using the data and correlative function. The concept of geographical information services in this paper is more extensive than generic geographical information services. They are some data services or some function services. The course for creating geographical information service ontology includes three phases. They are analyzing, conceptualization and realization. Their task is to illuminate the intention and the domain knowledge in the course of analyzing, to define the relation between conception and domain conception in the course of conceptualization, to code this domain conception model created in the course of conceptualization using formalization language in the course of realization, respectively. Also, its aim is to make the computer understand and access the conception model\(^3\). The course of the conceptualization is hard-core, and will save the result of conceptualization analysis through the felicitous form. There are four basic characteristics in the course of conceptualization: identity, integrality, severity and dependency. We can proceed with the four characteristics, clarify the class relation, identify the hierarchy and create a more reasonable ontology in the material analyzing course.

## 2 Multilevel matching arithmetic

Service matching is the basic question of geographical information service discovering. It aims to discover the services matching most the demand of service users from these services released. The service user matches to the service offers semantically in this course. The service matching arithmetic is very important to the entire technology of the geographical information service discovering. The basic function of service matching arithmetic is to select the accordant geographical information service from all these released services according to the user’s service request and output these compatible services\(^4,5\).

We consider that the service should match the request when the geographical information service has enough similitude to the service request. Then the key question is to define the condition of enough similitude. The request and the selected service are equal when they are the same geographical information service. However, the term of the same geographical information service is very rigorous. The service matching must be flexible. Importing to the credit similarity, we provide a new three-level service matching arithmetic through analyzing the service matching based on the semantic similarity\(^6,7\). The service aggregate becomes few in the course of the Web service discovering as shown in Fig.1.
2.1 Function similarity

It is important to compute a topologically correct and geometrically close terrain surface under certain conditions. OWL-S profile describes the input, output, prepositive precondition and result through hasInput, hasOutput, hasPrecondition and hasResult. Define the function similarity between the selected service (OBJ) and the service request (REQ) as Eq.(1):

\[
\text{sim}_{\text{fun}}(\text{REQ}, \text{OBJ}) = w_1 \text{sim}_{\text{input}}(\text{REQ}, \text{OBJ}) + w_2 \text{sim}_{\text{output}}(\text{REQ}, \text{OBJ}) + w_3 \text{sim}_{\text{pre}}(\text{REQ}, \text{OBJ}) + w_4 \text{sim}_{\text{effect}}(\text{REQ}, \text{OBJ})
\]

(1) Function similarity of OWL object

The function similarity between OWL object \( a \) and OWL object \( b \) is defined by their communal information quantity as Eq.(2):

\[
\text{common}(a, b) = \sum_{x \in \text{common}(a, b)} \text{IBIV}(x)
\]

\[
\text{desc}(a) = \{(a, p, o) \in O \}
\]

(3) Information value based on consequence

The information quantity included in the a member group of ontology is defined by its reasoning ability. The reasoning ability is defined by the member group number of new RDF created by some consequence rule. As Eq.(6):

\[
R_c(t, O) = \{ [b_1, b_2, \ldots, b_l] \mid \exists [a_1, a_2, \ldots, a_j] \in O \}
\]

(2) Description set

The communal information quantity of Eq.(1) is based on description. The description set desc(\( a \)) of OWL object \( a \) is defined by Eq.(3):

\[
\text{desc}(a) = \{(a, p, o) \in O \}
\]

Where \( O \) is the OWL object and also the object of ontology. \( p \) is the predication, \( a, p, o \) is the RDF triples.

For the description of the information quantity of OWL object ulteriorly, \( n \)-order description set is used as Eq.(4):

\[
\text{desc}_n(a) = \{(a, p, o) \in O \} \cup \text{desc}_{n-1}(x)
\]

(4) Information of the member group \( t \) converts to the inference information (inference-based information value, IBIV), the value is the number of the \( \text{inf}_s(t, O) \) element, as Eq.(9):

\[
\text{IBIV}(t, O) = \text{inf}_s(t, O) \cdot \text{IBIV}(t) = \text{inf}_s(t, O)
\]

According Eq.(9), define the \( f_{\text{common}}(a, b) \) and \( f_{\text{desc}}(a, b) \) of Eq.(2):

\[
f_{\text{desc}}(a, b) = \sum_{x \in \text{desc}(a, b)} \text{IBIV}(x)
\]

Import these two expressions to Eq.(2):

\[
\text{sim}(a, b) = \frac{\text{sim}_{\text{common}}(a, b) + \text{sim}_{\text{desc}}(a, b)}{\text{IBIV}(s)}
\]

Then we can use Eq.(10) to calculate the function similarity between the selected services and request.

2.2 Credit similarity

The semantic description of Web service is benefi-
cial to discover the geographical information service automatically. Also, the service discovering technology breaks away from the strict using only the traditional class and key word. The method can match faintly based on semantic description. Also, this method makes the service discovering technology more effective and veracious[9]. Sometimes, there are more than one geographical information services. How do we estimate the qualities of these services? Are these services all useable? How do we estimate the authenticity of these services? The credit similarity of geographical information service can resolve these questions. The key factors impacting credit similarity include response time, usability and expenses. The credit similarity can be defined as Eq.(11).

\[ \text{sim}_{\text{reputation}} = w_1 \text{sim}_r + w_2 \text{sim}_u + w_3 \text{sim}_p \]

\[ \sum w_i + w_2 + w_3 = 1 \] (11)

Where \( w_i \) is the similarity of the response time, \( \text{sim}_u \) is the similarity of usability, \( \text{sim}_p \) is the similarity of price.

(1) Similarity of the response time

Suppose \( t_1, t_2, \cdots, t_n \) is the register of the response time, average time can be calculated through the average time as Eq.(12).

\[ t = \frac{1}{n} \sum_{i=1}^{n} t_i \] (12)

Calculate the average time after eliminating the large error data using Eq.(12).

\[ \text{sim}_s = \begin{cases} 1, & t \leq t_{\text{request}} \\ \frac{t_{\text{request}}}{t}, & t > t_{\text{request}} \end{cases} \] (13)

The similarity of the response time can be defined as Eq.(13).

Where \( t_{\text{request}} \) is the threshold of response time.

(2) Similarity of usability

The similarity of usability is similar to the similarity of the response time.

(3) Similarity of price

The similarity of price is similar to the similarity of the response time.

We calculate the similarity by using the example of credit similarity. The credit similarity of geographical information service is one-dimensional array.

Suppose the similarity of the response time, usability and price: \( t_{\text{request}} = 0.01 \) second, \( a_{\text{request}} = 0.9 \), \( p_{\text{request}} = 100 \) Yuan. Suppose \( w_1 = 0.4, w_i = 0.35, w_3 = 0.25 \). The credit value matrix of a set of service (four selected services) will be:

\[
\begin{bmatrix}
0.02 & 0.017 & 0.009 & 0.003 \\
0.92 & 0.88 & 0.77 & 0.69 \\
101 & 97 & 91 & 110 \\
0.50 & 0.59 & 1.00 & 1.00 \\
1.00 & 0.98 & 0.85 & 0.76 \\
0.99 & 1.00 & 1.00 & 0.91
\end{bmatrix} \rightarrow
\begin{bmatrix}
0.50 & 0.59 & 1.00 & 1.00 \\
1.00 & 0.98 & 0.85 & 0.76 \\
0.99 & 1.00 & 1.00 & 0.91
\end{bmatrix}
\]

Function similarity:

\[
\begin{bmatrix}
0.50 & 0.59 & 1.00 & 1.00 \\
1.00 & 0.98 & 0.85 & 0.76 \\
0.99 & 1.00 & 1.00 & 0.91
\end{bmatrix} \times \begin{bmatrix}
0.4 & 0.35 & 0.25 \\
0.80 & 0.82 & 0.95 & 0.89
\end{bmatrix} = \begin{bmatrix}
0.80 & 0.82 & 0.95 & 0.89
\end{bmatrix}
\]

\[ 2.3 \text{ Multilevel matching arithmetic} \]

First, we match the general function documents between the request service and all the documents of serviceProfile, serviceMod, and ServiceGrounding in the enroll center in turn in the course of the document matching. Because these documents to be matched touch upon only some service description, there will be many services that accord to the request. We must filter some geographical information services thus maybe disturbing the result (some un-connected, inaccurate geographical information services). Then we put these services into the filter decider and decide whether to match these by the second-level matching. Put these services selected in the first-level matching into the register center if these services are congruent.

According to the similarity threshold, set beforehand, analyse the similarity between these services with the request services in function, and ascertain these services according to the request. Calculate the credit similarity from these services gained from the second-level matching in the third-level matching. Also, select the services according to the credit similarity.

3 Service discovering model

Through importing the OWL-S Profile information
to the data structure driven by UDDI, we can integrate the service ontology and UDDI technology organically. According to the point of view, carry out the OWL-S Profile information through using ontology to extend the UDDI. We provide the service register center driven by ontology based on the idea. The system structure is shown in Fig. 2.

The register center judges the message when the release ports receive the message of the released service. If the message does not include any semantic information, it deals with information according to the conventional UDDI fashion. If the message includes semantic information, it will deliver the message to the OWL-S UDDI conversion module. The OWL-S UDDI conversion module operates it according to the semantic information of the message. This means that the user can use the UDDI query interface to access the function of UDDI. However, the message cannot include any semantic information in this mode. In other words, the user cannot use the performance describing the OWL-S Profile information\[10,11\].

The user can not only use the current UDDI center to widen the bound of the geographical information service discovering, but also can query and release the geographical information service based on semantic information through enhancing the semantic information. The principle is that using the one-to-one mapping when the UDDI has the same elements as the OWL-S Profile through importing the OWL-S Profile information to the UDDI center such as the ContactInformation element of the OWL-S Profile. When the UDDI does not have the same elements as the OWL-S Profile, we will use the T-Model mapping. T-Model mapping is based on the loosely coupled mechanism between WSDL and UDDI. It is commended by the OASIS. It defines the special T-Model for every OWL-S Profile that does not have the mapping with UDDI, such as Input, Output, Service Parameter, etc.

**4 Discovering model test**

We have designed the archetype system to test the geographical information service discovering model. The system runs on a computer. The computer configuration includes an Intel Core2 processor (1.86GHz), 2GB EMS memory and Windows XP SP2 operating system. Because the aim of the test is to analyze and evaluate the model itself, all the servers run on the same computer in order to remove the influence of network transmission. Also, all the register information is registered in the same computer. Every geographical information service has a corresponding service description.

We select the Protégé ontology design software designed by Stanford Medical Informatics. We make two assumptions for the service register and service request before matching:

1. Suppose all the data and service provided by the service offer has the semantic description, and all the data and service is described by the OWL-S.
2. Suppose the service offer and service request are the different ontologies, and they consist of the same sharing ontology of the domain.

We test the geographical information service discovering model in terms of the recall ratio, precision, responsivity and load balance. The test environment includes domain environment (special domain, material domain and generic) and data environment. In the same test environment, note the conflict rate, recall ratio, and precision ratio of the service discovering result through creating the direct search discovering model, keyword discovering model and ontology-driven discovering model. Take samples from probability survey according to a definite period (one week, the response time and usability is according to
the run log). The test results are as following.

(1) Conflict rate of the service discovering

The conflict rate of the service discovering is the instance when selecting many optimization services in the course of the service matching and service discovering. In the course of service discovering, the more the semantic factors are, the more complicated the calculations are. Therefore, the run time will be extended. However, the result set will have the lower conflict rate. Contrarily, if the semantic factor is less, the result set will have the higher conflict rate. Fig. 3 and Fig. 4 illustrate the relation of the semantic factor, service number and time. The horizontal line delegates the number of the semantic factor, and the vertical line delegates the response time and the conflict rate. The data set includes 50, 100, 150, 200 and 250 services. Every node in the figure delegates that service set includes \( n \) services, the time that uses \( n \) semantic factor discover service. The value is defined as the statistical value\(^{[12,13]}\).

![Fig. 3 Relation between semantic factor and time](image)

**Fig. 3** Relation between semantic factor and time

According to Fig.3, the time will be linearly increasing along with the increment of the number of semantic factor. According to Fig.4, the infection of the number is very obvious when the number of semantic factor is less than four. The conflict rate is six percent when the number of semantic factor is ten. Four kinds of semantic factors are perfect when the services are less than 200. The number from four to ten is perfect when the services are more than 200.

(2) Recall ratio of the service discovering

We select 100 geographical services and corresponding ontologies to test the service discovering through the direct search discovering model, keyword discovering model and ontology-driven discovering model. The sampling probability and statistics result are shown in Fig. 5 (one week as the period). The recall ratio using the ontology-driven discovering model is higher than using keyword discovering model, especially in the material domain. Because the number of keywords in the material domain is few, the ontology-driven discovering model can ensure the recall ratio.

![Fig. 5 Discovering recall in different setting](image)

**Fig. 5** Discovering recall in different setting

Select different service set to test the geographical information service discovering model, the probability and statistics result are shown in Fig.6.

![Fig. 6 Discovering recall in different data setting](image)

**Fig. 6** Discovering recall in different data setting

(3) Precision ratio of the service discovering

The precision ratio is the percentage between the number accessed successfully and the number accessed when a user requests the service from the register center. In other words, the precision ratio is the percentage between the services discovered and all the services. The precision ratio is 96 percent using the ontology-driven discovering model (using one
week as a period in run log).

(4) Load of the service copy

The load of the service copy is the number of copy services having the same function being accessed\cite{14,15}. According to Fig. 7(a) and Fig. 7(b), take samples from probability and statistics of the copy services by using keyword discovering model and ontology-driven discovering model in a test period (one week). The result shows that the conventional discovering model does not control the load balance (Fig. 7(b)). Some copy services running may be in the upper bound. Some copy services may turn into the disabled state. The ontology-driven discovering model can control the load balance well.

![Fig. 7 Comparison between load of discovery ontology-driven and discovery based on traditional UDDI](image1)

(5) Response rate of the service discovering

The response rate of service discovering is the ratio between the number getting one service at lowest and the whole request\cite{16}. Fig. 8 is the sample result in the running period of the system. According to Fig. 8, the response rate of the traditional UDDI model is in a downtrend. The response rate of ontology-driven discovering model is clearly higher than the traditional UDDI model.

![Fig. 8 Recall of the geographical information service requests](image2)

5 Conclusion

The ontology-driven discovering model is based on the geographical information service ontology and service semantic description. According to the determinative threshold, the final suited geographical information services are obtained by analyzing the function similarity and the credit similarity between the services to be selected and requested services, filtering and matching the services to be chosen in the register center. We have tested the ontology-driven discovering model. The result indicates that the ontology-driven discovering model is excellent in recall ratio and precision ratio, and also can maintain the dynamic load balance of the service copy.

In ontology-driven discovering, the design of geographical information service and semantic description is very complicated. We have designed only the test environment of the ontology-driven discovering
model. In application, the number of the geographical information services and the environment would affect the service discovering. Also, the similarity threshold currently has no uniform standard for the actual application.

Reference

[1] Li Deren, Cui Wei (2004) Semantic grid of spatial information[J]. Geomatics and Information Science of Wuhan University, 10(29): 847-851
[2] Cui Wei, Li Deren (2005) Architecture of resource management of spatial information grid based on ontology and LDAP [J]. Geomatics and Information Science of Wuhan University, 30(6):549-552
[3] Li Deren, Cui Wei (2006) Geographical ontology and special information multi-level gridding[J]. Acta Geodaetica et Cartographica Sinca, 35(2): 143-147
[4] Hang Yu (2006) Research on semantic web service based on ontology [D]. Taiyuan:Taiyuan University of Technology
[5] Chen Wei(2006)Research on pivotal technology of semantic web services [D]. Xi’an: Northwestern Polytechnical University
[6] Paolucci M, Kawamura T, Payne T (2002) Semantic matching of Web services capabilities[C]. Proc. of the International Semantic Web Conference 2002, Sardinia
[7] Martin D, Paolucci M, McIlraith S, et al. (2004) Bringing semantics to web services: the OWL-S approach [C]. SWSWPC 2004, San Diego, CA
[8] Shi Lei (2006) Research on the semantic Web service discovering technology [D]. Nanjing: Hehai University
[9] Chen Changsong (2002) A preliminary study of the data-sharing-oriented GIS semantic representation theory[J]. Acta Geodaetica et Cartographica Sinca, 29(1): 193
[10] An Yang, Bian Fuling (2004) The description and discovering of net geographical information service based on ontology[J]. Geomatics and Information Science of Wuhan University, 29(12): 1063-1066
[11] Klein M, Bernstein A (2001) Searching for services on the semantic Web using process ontologies[C]. Proc. of the International Semantic Web Working Symposium, Stanford
[12] Li Xiang, Zhang Xihui, Lin Hui (2009) Deriving network-constrained trajectories from sporadic tracking points collected in location-based services[J]. Geo-spatial Information Science, 12(2): 85-94
[13] Lian Huiqun (2005) The application and research on the combination and discovering of semantic Web service [D]. Fouzhou :Fouzhou University
[14] An Yang (2005) The research on the key question of network geographical information based on ontology[D]. Wuhan: Wuhan University
[15] Caprotti O, Dewar M, Turi D (2004) Mathematical service matching using description logic and OWL[C]. Proc of the Mathematical Knowledge Management, Phoenix, Arizona
[16] Bussler C, Fensel D, Maedche A (2002) A conceptual architecture for semantic Web enabled Web services[J]. SIGMOD Record, 31(4): 24-29