Journal of Southeast University (English Edition) Vol. 22, No. 3, pp. 297 - 301 Sept. 2006 ISSN 1003—7985

Web services selection based on semantic similarity

Zhu Junwu[1, 2] Wang Jiandong[1] Li Bin[2, 3]

([1] School of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)
([2] School of Information Engineering, Yangzhou University, Yangzhou 225009, China)
([3] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

Abstract: Services discovery based on syntactic matching cannot adapt to the open and dynamic environment of the web. To select the proper one from the web services candidate set provided by syntactic matching, a service selection method based on semantic similarity is proposed. First, this method defines a web services ontology including QoS and context as semantic supporting, which also provides a set of terms to describe the interfaces of web services. Secondly, the similarity degree of two web services is evaluated by computing the semantic distances of those terms used to describe interfaces. Compared with existing methods, interfaces of web services can be interpreted under ontology, because it provides a formal and semantic specification of conceptualization. Meanwhile, efficiency and accuracy of services selection are improved.

Key words: web services; ontology; semantics; similarity

Web service, as a new model of web application and distributed computing, is an effective mechanism for the integration of data and applications on the web[1−2]. As a consequence, it develops quickly. First, the quantity of web services is increasing constantly and their functions are also changing ceaselessly[3]. Secondly, web services and their providers are globally distributed[4]. Unfortunately, users and machines have no ways to understand and process massive, dynamic and hetero-structured services, because the existing services cannot be supported by formal semantics[5−7].

In order to select proper one from amount of web services to satisfy certain requirements, a new technique is needed to describe and organize web services. Meanwhile, a mechanism to compute the degree of satisfiability is also important for web service selection from a candidate set.

Aiming at the selection of web services, this article presents an overview of the method of web selection based on semantic similarity.

1 Related Work

Web services selection is the key link of service discovery, and is also a pre-condition for service composition. So, it has become a hotspot in the research domain of web services.

Web services discovery based on syntax is a very effective method of providing a web services candidate set. But with the increase of the number of developers and users of web services and the overlapping of multiple subjects and domains, the unsharable vocabulary will result in the failure in efficiency and accuracy. Especially for the second selecting from a candidate set, syntactic matching is almost hopeless[8]. The BALES method[9] is a typical web services selection method based on semantic matching. It uses the ontology of WordNet to gain the sharable explanations in the domain, and is also a sharable glossary of descriptions on service.

DAML-S provides a top-level ontology to describe services and includes the specifications of function and quality limit[10]. Concerning QoS, IBM presents WSLA (web service level agreements), and it is SLA instruction based on XML[11−12]. WSOL (web service offerings language) also supports the format instruction for every limit[13]. The web services framework which supports QoS is first presented in Refs. [11, 14]. It supports the validation on QoS, but does not give a good definition of the framework, and does not give the validation process in detail. Chen et al. [15] presented a service discovery model, and this model includes the function requirements and the non-function requirement (such as quality).

CB-SeC[16], an agent-based, framework, computes and classifies web services according to the information and service environment of consumers. Aura framework[17] models on users' task, and users' task adjusts with the change of environment. At the same

Received 2006-04-05.

Foundation items: The National Natural Science Foundation of China (No. 70471090, 70472005), the Natural Science Foundation of Jiangsu Province (No. BK2004052, BK2005046).

Biographies: Zhu Junwu (1972—), male, graduate, jwzhu@ yzu. edu. cn; Wang Jiandong (1945—), male, professor, aics@ nuaa. edu. cn.

time, when the position of user changes, Aura system will finish the transfer from primary environment to new environment. Ref. [18] used context WIDEGET to collect context information from sensors, then aggregate and explain information.

2 Web Services Ontology

Ontology is a formal explicit specification of concepts in a domain of discourse, properties of each concept describing various features and attributes of the concept, and restrictions on slots[19-20]. Concepts denote sets of any individuals, and roles denote binary relationships between individuals. The extension of any concept is a subset of any individuals set, while the extension of any role is a subset of ordered pairs of individuals. In this sense, we define individuals as a component of terminology.

Definition 1 In web services domain, terminology T is a 3-tuple:

$$T = \{C, R, I\} \tag{1}$$

where C and R are concepts set and roles set, respectively, and I is individuals set. If we use $V(O)$ to express the glossary of O; $V_C(O)$, $V_R(O)$, $V_I(O)$ respectively express the concepts set, roles set, and individuals set of the ontology O, then

$$V(O) = V_C(O) \cup V_R(O) \cup V_I(O) \tag{2}$$

T provides a set of terms to describe interfaces of semantic web services.

Definition 2 One semantic web service (SWS) can be described as a 3-tuple:

$$\mathrm{SWS}(M_{IN}, M_{OUT}, O)$$

where $M_{IN} \subseteq V(O)$; $M_{OUT} \subseteq V(O)$; SWS is the name of the semantic web service. M_{IN} and M_{OUT} are input interfaces sets and output interfaces sets of SWS; O is the ontology which supports SWS, and the concepts and roles that the input sets and output sets of SWS used are from ontology O.

Definition 3 Let T be a terminology of certain domain, and then domain ontology is a partial ordered relation on T.

$$O = \langle T, \subseteq \rangle \tag{3}$$

Not only definitions in T but also axioms in ontology can be transformed as inclusion relations, and the hierarchy of all concepts, roles and individuals, organized by ISA (or AKO), can be translated into inclusion relations. So, ontology is denoted as Poset (partial ordered set).

Quality and context are the important properties of service selection. For the consumers of web service, they must consider the QoS factor including usability, security, respond time and throughput, etc. To valuate web services, a context must be donated, because the performance of web service is closely related to it. Context of web services means the application conditions and environments. An agent apperceives conditions, and executes relevant actions according to environments. For the agent of service consumer, context is all the information about client of web services. The information can be used to adjust executing and output, and its ultimate purpose is to satisfy user customization and individual requirements. Fig. 1 is the structure of web services ontology including QoS and context.

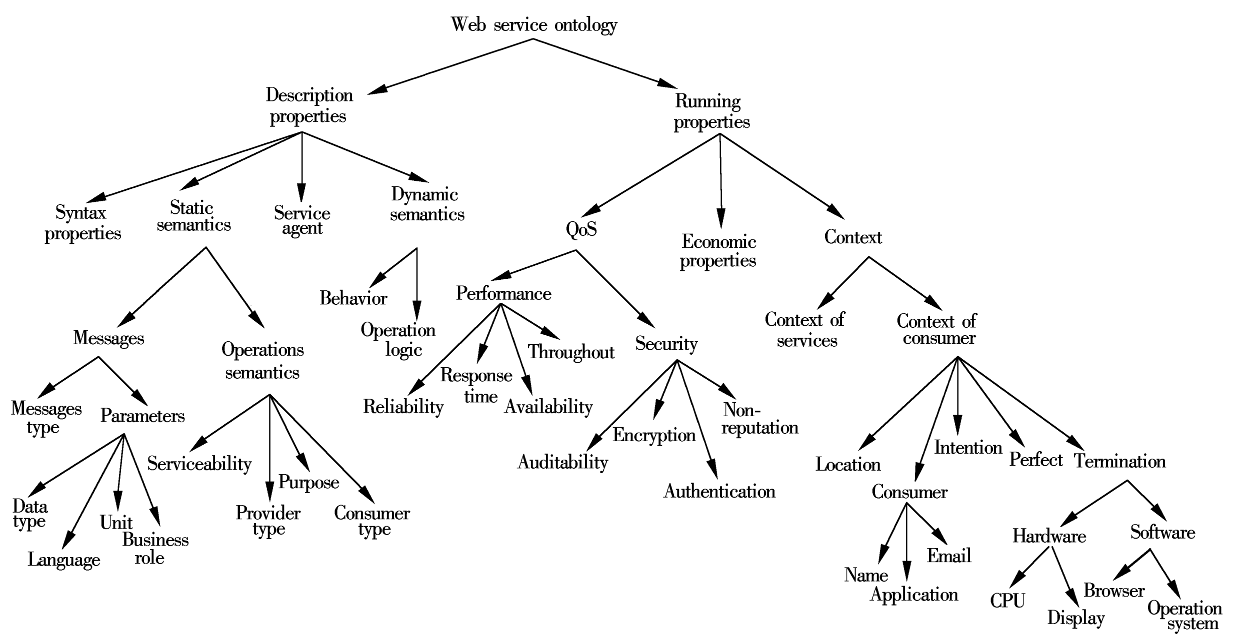

Fig. 1 Web services ontology including QoS and context

3 Computation of Semantic Degree

When concept a is an interface of web service A and b is an interface of web service B, we use similarity cell to describe them, denoted as $u(a, b)$.

One concept, as a term, may have many sub-properties, and may be denoted as $s_1, s_2, \ldots, s_m$ in turn. $E_j(a)$ is the eigenvalue of s_j of super-property a. Eigenvalue proportion of interfaces a and b relative to sub-property s_j is denoted as

$$r_{\frac{a}{b}} = \frac{\min\{E_j(a), E_j(b)\}}{\max\{E_j(a), E_j(b)\}} \tag{4}$$

where $j = 1, 2, \ldots, m; 0 < r_{\frac{a}{b}} \leqslant 1$.

According to the effect of similarity degree for sub-properties $s_1, s_2, \ldots, s_m$ to a and b, we present the powers $d_1, d_2, \ldots, d_m$ of $s_1, s_2, \ldots, s_m$, respectively. d_j means the power value of effect on similarity degree for the j-th sub-property to similarity degree of interfaces, then the value of the i-th similarity cell $u_i(a, b)$ between web services A and B is denoted as

$$q(u_i) = d_1 r_1 + d_2 r_2 + \ldots + d_m r_m = \sum_{j=1}^{m} d_j r_j \tag{5}$$

where $i = 1, 2, \ldots, n; 0 < r_j \leqslant 1; 0 \leqslant d_j \leqslant 1; \sum_{j=1}^{m} d_j = 1$. There are many methods of giving power, such as symbolic statistics, fuzzy mathematics, experimentation or expert wisdom. When $q(u_i) = 0$, it means not any similarity between two elements; when $0 < q(u_i) < 1$, it means the two elements are similar, the value of similarity cell reflects the size of similarity degree; when $q(u_i) = 1$, it means all sub-properties of two concepts are equivalent.

Suppose that web service A consists of k interfaces, web service B consists of l interface, n similarity elements between system A and B compose n similarity cells. The value of similarity cell is denoted as $q(u_i)$, w_i is the power value of the effect on similarity degree for every similarity cell of two web services. Their similarity degree is denoted as

$$Q = f(k, l, n, m, q(u_i)) = \frac{n}{k + l - n} \sum_{i=1}^{n} w_i q(u_i) \tag{6}$$

where $i = 1, 2, \ldots, n; 1 \leqslant n \leqslant \min(k, l); k = 1, 2, \ldots, N; l = 1, 2, \ldots, N; 0 \leqslant w_i \leqslant 1; \sum_{i=1}^{n} w_i = 1$. According to the similarity degree, web services are classified by homology web services: $k = l = n, q(u_i) = 1, Q = 1$; similar web services: $0 < Q < 1$, k, l, n are not congruent, and $q(u_i) \neq 0, q(u_i) \neq 1$; while dissimilarity web services: $Q = 0, n = 0$.

According to the similarity principle, the basic process of the semantic similarity degree algorithm is to confirm similarity elements, construct similarity cells, and compute proportion of sub-properties, powers of characters, similarity degree and so on[21]. Concrete steps are as follows:

1) Confirming similarity elements

Ontology can be described by many methods such as natural language, ontology description language, formalization, semi-formalization, frame description and so on. Whatever method is used, one concept or class includes attributes, relations, axioms and objects. $V_C(O)$, $V_R(O)$ and $V_I(O)$ can be described as a intersection of many sub-properties. The attributes or eigenvalues of interfaces reflect the semantic similarity degrees of concepts.

We need to distinguish essential attributes of one interface, because it provides more contributions to compute similarity degree. We also need to distinguish the steady characters from the changeable.

2) Constructing similarity cells

Sub-properties of interfaces are the similarity cells in our method. For concepts and objects, they may be classified into three types, M levels similarity cells.

Due to limitations of space, in the following depictions, we just take attributes for an example, suppose two concepts A and B, in the first level $U_C = (V_C(O_A), O_C(O_B))$; in the second level $u_{C_i} = (a_i, b_i)$, wherein $a_i \in O_C(O_A)$, $b_i \in O_C(O_B)$, $1 \leqslant i \leqslant n$ (see Fig. 2).

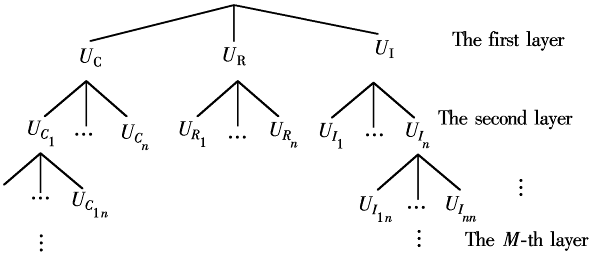

Fig. 2 Similarity cell structure of concepts

3) Defining the value of elements in the M-th layer (bottom)

Select the eigenvalues of elements according to the attributes in the M-th level.

$$S_C = \{S_{C_1}, S_{C_2}, \ldots, S_{C_n}\} \tag{7}$$

For computing the eigenvalues of ontology concepts we can describe them by using ontology directed graph to compute semantic distance.

4) Computing the value of the M-th level similarity cells

$$q_C(u_{C_i}) = \sum_{j=1}^{m} d_{C_j} r_{C_{ij}} \quad (8)$$

where d_{C_j} is the power of the M-th similarity elements, $r_{C_{ij}}$ can be computed by using Eq. (4).

5) Computing eigenvalues and the value of similarity cells at every level

Repeat 3) and 4), compute the eigenvalue and the value of similarity cells at every level from the ($M-1$)-th level to the first level.

6) Computing similarity degree of web services

Use Eq. (6) to compute semantic similarity degree, w_i is the similarity power value of the first level, the domain of Q is $[0,1]$. $Q=0$ means that the semantic similarity degree of concepts or objects is 0, is not completely. $Q=1$ means they are equal. If $Q \in (0,1)$, the closer Q inclines to 1, and the greater the similarity degree is.

4 Conclusion

Aiming at services selection under the open and dynamic environment, a set of terms to describe web services has been defined, and this vocabulary can be used as a template of semantic interpretation. By computing the semantic distance of web services interfaces, we can evaluate if the selected web services can satisfy the consumer requirements.

References

[1] Berners-Lee T, Fischetti M, Dertouzos T M. *Weaving the web: the original design and ultimate destiny of the world wide web by its inventor*[M]. San Francisco: Harper Collins Publishers, 2000. 1-256.

[2] Kim J B, Segev A. A web services-enabled marketplace architecture for negotiation process management [J]. *Decision Support System*, 2005, **40**(1): 71-87.

[3] Medjahed Brahim. Semantic web enabled composition of web services[D]. Virginia, USA: ProQuest Information and Learning Company Church, 2004. 1-279.

[4] Sycara K. Dynamic service matchmaking among agents in open information environments[J]. *ACM Sigmod Record*, 1999, **28**(1): 47-53.

[5] McIlraith S. Modeling and programming devices and web agents[A]. In: *NASA Goddard Workshop Formal Approaches to Agent-Based Systems, Lecture Notes in Computer Science*[C]. Springer-Verlag, 2001. 63-67.

[6] Hendler. Agents and the semantic web[J]. *IEEE Intelligent Systems*, 2001, **16**(2): 30-37.

[7] Martin D L, Cheyer A J, Moran D B. The open agent architecture: a framework for building distributed software systems[J]. *Applied Artificial Intelligence*, 1999, **13**(1,2): 91-128.

[8] Davies J, Weeks R. QuizRDF: search technology for the semantic web[A]. In: *Proceedings of the 37th Annual International Conference*[C]. Hawaii, 2004. 112-119.

[9] van den Heuvel W J. Integrating modern business applications with legacy systems [D]. Tilburg University, 2002.

[10] Miller G A. WordNet: a lexical database for English [J]. *Comm of the ACM*, 1995, **38**(11): 39-41.

[11] Tang Q C, Cheng H K. Optimal location and pricing of web services intermediary [J]. *Decision Support System*, 2005, **40**(1): 129-141.

[12] Keller A, Ludwing H. The WSLA framework: specifying and monitoring service level agreements for web services [J]. *Journal of Network and Systems Management*, 2003, **11**(1): 57-81.

[13] Tosic V, Pagurek B, Patel K. WSOL a language for the formal specification of classes of service for web services [A]. In: *International Conference on Web Services* [C]. Las Vegas, Nevada, USA, 2003. 375-381.

[14] Tian M, Gramm A, Naumowicz T, et al. A concept for QoS integration in web services [A]. In: *The 4th International Conference on Web Information Systems Engineering* [C]. Rome, Italy, 2003. 149-155.

[15] Chen Hongan, Yu Tao, Lin Kwei-Jay. QCWS: an implementation of QoS-capable multimedia web services [A]. In: *IEEE the Fifth International Symposium on Multimedia Software Engineering* [C]. Taichung, 2003. 38-45.

[16] Peddemors A J H, Lankhorst M M, de Heer J. Presence, location, and instant messaging in a context-aware application framework [A]. In: *Proc of the Intl Conf on Mobile Data Management (MDM)*, *LNCS*[C]. Springer-Verlag, Melbourne, Australia, 2003, **2574**: 325-330.

[17] Sousa J P, Garlan D. Aura: an architectural framework for user mobility in ubiquitous computing environments [A]. In: *Proc of the Working IEEE/IFIP Conf on Software Architecture (WICSA)* [C]. Washington DC, 2002. 29-43.

[18] Dey A K, Salber D, Abowd G D. A context-based infrastructure for smart environments [A]. In: *Proc of the Intl Workshop on Managing Interactions in Smart Environments (MANSE)* [C]. Dubin, Ireland, 1999. 114-128.

[19] Duineveld A J, Stoter R, Weiden M R, et al. WonderTools? A comparative study of ontological engineering tools [J]. *International Journal of Human-Computer Studies*, 2000, **52**(6): 1111-1133.

[20] Uschold M, Gruninger M. Ontologies: principles, methods and applications [J]. *Knowledge Engineering Review*, 1996, **11**(2): 93-136.

[21] Zhou Limei. *Similarity principle* [M]. Beijing: Chinese Science and Technology Press, 1993. 1-322. (in Chinese)

基于语义相似性的 web 服务选择

朱俊武[1,2] 王建东[1] 李 斌[2,3]

([1] 南京航空航天大学信息科学与技术学院, 南京 210016)
([2] 扬州大学信息工程学院, 扬州 225009)
([3]南京大学计算机软件新技术国家重点实验室,南京 210093)

摘要:基于语法匹配的服务发现无法适应 web 的开放性和动态性环境. 为了对语法匹配产生的 web 服务候选集进行二次选择,提出一种基于语义相似性的 web 服务选择方法. 该方法首先定义了包含 QoS 和上下文的 web 服务本体. 服务本体提供了一个接口描述的通用术语集合,以此作为服务描述的语义支撑. 然后,通过相似理论和 web 服务本体计算 web 服务接口之间的语义距离. 与已有的方法相比,由于提供了概念化的形式语义说明,web 服务的接口能在本体下得到准确的解释,同时服务选择的效率和准确性得到了改进.

关键词:web 服务;本体;语义;相似性
中图分类号:TP391. 1