E-ISSN: 2395-1702 P-ISSN: 2395-0382 Volume 03- Issue 06- pp-08-13

**Research Paper** 

Open Access

### ONTOLOGY BASED-GENERALIZED ASSOCIATION RULE ALGORITHM FOR WEB SERVICE COMPOSITION AND DISCOVERY IN SERVICE REPOSITORIES

## Manikandan.A<sup>1</sup>, Danapaquiame.N<sup>2</sup>, Govindharaj.I<sup>3</sup>

Associate Professor<sup>1</sup>,PG Scholar<sup>2</sup>, Assistant professor<sup>3</sup> Department of CSE, Sri ManakulaVinayagar Engineering College, Pondicherry, India<sup>1,2,3</sup> <u>maniraghu22@gmail.com<sup>1</sup>,n.danapaquiame@gmail.com<sup>2</sup>,gvraj87@gmail.com<sup>3</sup></u>

#### **Abstract:**

The number of Web services has grown drastically. Then how to manage them efficiently in a service repository is an important issue to address. Given a special field, there often exists an efficient data structure for a class of objects, e.g., the Google' Big table is very suitable for Web pages' storage and management. Based on the theory of the equivalence relations and quotient sets, this work proposes a multilevel index model for large-scale service repositories, which can be used to reduce the execution time of service discovery and composition. Its novel use of keys as inspired by the key in relational database can effectively remove the redundancy of the commonly-used inverted index. Its four function-based operations are for the first time proposed to manage and maintain services in a repository. The experiments validate that the proposed model is more efficient than the existing structures, i.e., sequential and inverted index ones. **Keywords:** Web service, service composition, service discovery, service management, and ontology

#### 1. Introduction

Data mining, the extraction of concealed prescient data from huge databases, is an effective new innovation with extraordinary potential to help organizations concentrate on the most critical data in their information distribution centers. Information mining devices foresee future patterns and practices, enabling organizations to make proactive, learning driven choices. The mechanized, forthcoming examinations offered bv information mining move past the investigations of past occasions given by review devices run of the mill of choice emotionally supportive networks. Information mining instruments can answer business addresses that customarily were excessively tedious, making it impossible to

determine. They scour databases for concealed examples, finding prescient data that specialists may miss since it lies outside their expectations. Most organizations effectively gather and refine monstrous amounts of information.

#### 1.1 Data Mining

Data mining is the abstraction of hidden data from a large data sets which is used to develop the business readiness. Expanded availability of huge amounts of data are possible for turn such information and knowledge into a useful and understandable data. The data mining is also has the development of data cleaning, pattern evaluation, data presentation and data transformation. It plays an important role in identifying the objects covered data patterns

and produce the detailed information in the business. However, precisely because XML is a kind of semi-structured text data, it has many weaknesses as text documents and semi -structured data. like, analysis the document must read it in the order form, the visit is inefficient; irregular organize the information, or may be constantly changing its structure or even may be incomplete, etc. The traditional data mining technology is the main face to the structured data-based relational database, transaction database and data warehouse [4]. Thus we cannot apply the traditional relational data -based mining methods, such as Apriori, to the semistructured data mining directly. Therefore, to develop effective methods for XML data mining become an important issue in the field of data mining and XML technology research Information areas. mining procedures can be executed quickly on existing programming and equipment stages to improve the benefit of existing data assets, and can be incorporated with new items and frameworks as they are brought on-line. At the point when actualized on superior customer server or parallel preparing PCs, information mining devices can dissect monstrous databases to convey. Data mining in includes to "removing or mining" learning from a lot of information. There are numerous different terms conveying a comparative or marginally unique intending to information mining, for example, learning mining from databases, information extraction, information/design examination. information archaic exploration, and information digging. Many individuals regard information digging as an equivalent word for another prominently utilized term, Knowledge Discovery in Databases", or KDD. On the other hand. others see information mining as basically a basic stride during the time spent learning revelation in databases.

There has been a number of web service composition methods proposed bv researchers. Some of the web service composition methods are reviewed as follows: Tang et al. [1] introduced a novel automatic Web service composition method based on logical inference of Horn clauses and Petri nets. They first transform a Web service composition problem into a logical inference problem of Horn clauses based on the forward-chaining algorithm. They then use the Petri net and its structural analysis techniques to obtain the composite service. Since there may be a large number of services in a service repository, and a huge number of rules may be generated consequently, the Petri net of a Horn clause set is very large. In order to reduce the composition time, they proposed a method to select the candidate clauses for the inference when a new query comes. Its weakness is that it must be executed after receiving user requirements, and cannot be executed beforehand and the optimization process cannot be executed before receiving requirements.Wu user and Khoury [2]proposed a tree-based search algorithm for Web service composition in a cloud computing platform. They first create a tree that represents all possible composition solutions according to user requirements, and then prune the illegal branches aiming to reduce response time and improve performance, and finally use a heuristic algorithm to search an optimal solution. This method has the disadvantage similar to that in [1] ie the optimization process cannot be executed before receiving user requirements. Talantikite et al[3] presented a Web Services model automatic for Discovery and its Composition. In order to understandable descriptions. Semantic Annotation is used for web service Discovery and composition. The proposed approach uses from an inter-connected network of semantic Web services

# **Related Work**

describing in OWL-S using the similarity measure between concepts like pellet before any submitted request. Their proposed approach gives several composition types: serial, dependent parallel and independent parallel. The Semantic Network is explored in backward chaining and depth-first in a single pass. At the end, are obtained several composition plans that satisfy the request and only one optimal composition plan using QoS is returned to the requester.

Wang and Guttula et al [4] presented a semiautomatic approach for web service composition that including both data mediation and service suggestion algorithms. This paper seeks to aid users trying to compose web services into a process by providing service suggestions. A graph IODAG (Input Output Directed Acyclic Graph) is defined to formalize an input/output schema of a Web service operation. Three data mediation algorithms leaf-based, structure-based and pathbased developed address are to data heterogeneities in process design. For adding semantic description into web used from services have Semantic Annotation for WSDL and XML Schema. This approach utilized various types of annotations and QoS. Finally they have developed a data mediation approach that tries to find automatically the optimal mappings between outputs and inputs. And finally came to the conclusion that pathbased algorithm is best data mediation algorithm from other two algorithms. Lee et al. [5] proposed scalable and efficient Web service composition method based on a relational database. They also uses the service net as a basic data structure. The service net has two drawbacks. First, it does not consider the issue to facilitate service discovery. Second, it is time-consuming for service addition and deletion [5]. The behavioral description-based Web Service Composition (WSC)problem deals with the

automatic construction of a coordinator web service that controls a set of web services to reach the goal states. Despite its importance and implications, very few studies exist on the computational complexities of the WSC problem.D. Lee et al[6] proposed a Treebased search algorithm for web service composition in SaaS With the growing demand of cloud computing most companies are moving software to the cloud in form of web services. Web services are popular in terms of distributed technology that can successfully solve integration problems between heterogeneous systems. Web service composition is NP-hard problem and one of the most challenging problems in services.J. web Kwon et al[7] proposedRedundant-Free Web Services Composition Based on a Two-Phase Algorithm.In this paper they proposed a redundant-free web services composition search based on a two phase algorithm. In the forward phase. the candidate composition will be found efficiently by searching the Link Index. In the backward phase. redundant-free web services compositions are generated from the candidate composition by using the concepts of tokens.D. Lee et al [8] proposed a computational complexity of behavioral description-based web service composition. This behavioural description-based Web Service Composition (WSC) problem deals with the automatic construction of a coordinator web service that controls a set of web services to reach the goal states. Despite its importance and implications, very few studies exist on the computational complexities of the WSC problem.

### 3. ALGORITHM DESCRIPTION 3.1ONTOLOGY

Ontology is a representation of knowledge formal. It provide a clear and consistent representation of terminology and methods that help people to observe the problems and dealing with affairs, provide public vocabulary of areas and define different levels of formal meanings of terms and relationships between terms. It is organized by taxonomy, and includes the typical model of the original language of the ontology and can provide a public and consistent understanding of the field. It overcomes the semantic content of the communication mismatch problem.

#### 3.2 ONTOLOGY BASED WEB MINING

Internet has become an indispensable tool usage for everyone, Web mining correspondingly becomes a hotspot, which uses huge amounts of data in the web server and other relevant datasets for mining analysis and gains valuable knowledge model about usage of relevant Web site. Nagi et al describe a work utilizes association rule mining integrated with fuzziness factor in order to analyze weblog data. The target is to find pages that are accessed together by majority of the users and hence should be linked in a proper way in order to maximize user satisfaction by providing to the users access flow. This way the number of visitors to the analyzed website will be maximized and hence the target will be achieved. Zhu et. al. [184] propose a new vector space retrieval algorithm based on association diagram extension of key words. By using key words and the related words appearing Ontology structure is divided into the following five stages :

a) Identify the purpose and scope of the ontology application: establish the field of study: establish the corresponding domain ontology or process ontology.

b) Ontology analysis: define the relationship between all terms and ontology meaning.

c) Represent ontology: to select a proper method of ontology according to the system need.

d) Ontology test: main test the clarity, consistency, integrity, scalability of ontology.

e) Ontology building: test the ontology according to the above criteria, to meet the requirements store the file form, otherwise switch to

Ontology representation: In order to describe and represent ontology, in recent years appeared a variety of ontology language. This paper choose one of the languages----OWL ontology (Web Ontology Language) [15]. OWL is designed by the World Wide Web Consortium Web Ontology Working Group. Its syntax is very similar to DAML + OIL, and can easily be converted to the latter. OWL can be used to clearly express the meaning of the vocabulary entries as well as the relationship between these entries. This express of vocabulary entries and relationship between them is called Ontology. OWL has more mechanism to express semantics than XML

RDF and RDF Schema, so it exceed capability of XML, RDF and RDF Schema that can only expressed machine-readable document content online

#### **3.2ONTOLOGY-BASED**

# ASSOCIATION RULES XML MINING APPROACH

In this paper, the traditional algorithm is as the basis to improve the proposed algorithm. Mining object of traditional algorithm is transaction-oriented database, the Ontology-based XML mining association rules mining algorithm is an object-oriented XML data source. In this article, the XML data source stored in to the hash table during the data pre-processing time to speed up the speed of traversing the entire XML tree. Meanwhile, the algorithm in this paper does three improvements on the traditional Apriori algorithm: First, when stored the XML data source in to the hash table and to each transaction and each data item is also added to all the parent of this transaction, then add the parent to candidate item sets as normal data, when traversing XML trees, also count on the parent class.

Second, according to the pre-existence table hierarchy of domain ontology, Can get parent of each item sets and remove the parent class does not appear in any concentration of the candidate; Third, pruning also include key items of the item *x* and its ancestor set.

# 3.3 INFORMATION RETRIEVAL MINING

Information extraction refers to the techniques designed to identify useful information from text documents It has the goal automatically. of transforming a collection of documents, usually with the help of an IR system, into information that is more readily digested and analyzed. By converging the semantic web and web mining techniques, e-Healthcare evolved a consumer-centered model of health care where various stakeholders collaborate utilizing information and technologies communication ICTs) including Internet technologies to manage and care for individual's health. Following are the important factors that equip patients with the essential knowledge for managing and maintaining better health.

### 4. CONCLUSION

This work proposes a multilevel index model to store and manage services for large-scale service repositories. Based on the theory of the equivalence relations and quotient sets, four level indexes are given to construct the multilevel index model. Four operations are proposed to manage and maintain services based on functions implied in the proposed model. The theoretical analysis and experimental results validate that the proposed multilevel index is more efficient and stable for service discovery and composition than the sequential structure and inverted index. Especially, the advantages of the proposed model become clearer and more significant as the number of service increases. Our experiments validate that our four operations

are efficient. In the era of drastically expanding services, the proposed model provides a highly desired storage structure for large-scale service repositories.

#### **5. REFERENCES**

[1] Y. Wu, C. Yan, Z. Ding, G. Liu, P. Wang, C. Jiang, et al., "A Novel Method for Calculating Service Reputation," IEEE Transactions on Automation Science and Engineering, vol. 10, pp. 634-642, 2013.

[2] C.-S. Wu and I. Khoury, "Tree-based Search Algorithm for Web Service Composition in SaaS," in Proc. the Ninth International Conference on Information Technology: New Generations (ITNG), 2012, pp. 132-138.

[4]S. N. Han, G. M. Lee, and N. Crespi,
"Semantic Context-Aware Service Composition for Building Automation System," Industrial Informatics, IEEE Transactions on, vol. 10, pp. 752-761, 2014.
[5] S. C. Oh, D. Lee, and S. R. Kumara, "Effective Web Service Composition in Diverse and Large-Scale Service Networks," IEEE Transactions on Services Computing, vol. 1, pp. 15-32, 2008.

[6] D. Lee, J. Kwon, S. Lee, S. Park, and B. Hong, "Scalable and efficient web services composition based on a relational database," Journal of Systems and Software, vol. 84, pp. 2139-2155, Dec. 2011.

[7] J. Kwon, H. Kim, D. Lee, and S. Lee, "Redundant-Free Web Services Composition Based on a Two-Phase Algorithm," in Proc. the 2008 IEEE International Conference on Web Services, 2008, pp. 361-368.

[8]W. Nam, H. Kil, and D. Lee, "On the computational complexity of behavioral description-based web service composition," Theoretical Computer Science, vol. 412, pp. 6736-6749, 2011.

[9] S. C. Oh, D. Lee, and S. R. T. Kumara, "Web service planner (WSPR): An effective and scalable web service composition

2017

algorithm," International Journal of Web Services Research, vol. 4, pp. 1-22, Jan-Mar 2007.

[10] R. Hewett, P. Kijsanayothin, and B. Nguyen, "Scalable Optimized Composition of Web Services with Complexity Analysis," in Proc. the 2009 IEEE International Conference on Web Services, 2009, pp. 389-396..