Abstract
The current internet has seen a tremendous growth of web services as an important technology for exchanging information, computing resources and programs online. With these increasing acceptance and existence of internet services, it has become high importance to have effective and accurate recommendation systems. It was observed that vast majority of web services are without associated with a precise description. Due to which most relevant service are not discovered to a particular user service request. Semantic based service discovery are mostly considered for service classification in the similar activities services. But these approaches are limited in their classification to the domain trained knowledge for the service discovery. This paper propose a novel self-adaptive semantic classification approach using service knowledge ontology and web user log frequent pattern in combination to improvise the web service discovery. The performance evaluation measures shows an improvisation in the classifying ability utilizing the service information of a service specific knowledge and user web usage patterns.

Keywords: Web Service discovery, Semantic Classification, Usage Mining, Frequent Pattern, service ontology

Introduction
In a service-oriented architecture web services system are to facilitate the creation of a large number of distributed applications over the Web [1],[9],[10]. It offers variety of activities in different areas of the Internet services and they are released and invoked for various service requirements. Web service discovery groups the services through semantic which are similar in services. However, as the service descriptions are syntactical in nature it is challenging to suggest the appropriate service selection and matching process for the different user request. Existing approaches for the service discovery are mostly based on the key-matching techniques to locate web services which are published [2],[3]. This discovery approaches which are based on the syntax-based matchmaking for the request does not provide accurate results. As a result, the service request can be considered as the choice of a few services that match with precise syntactical. Thus, to selects appropriate service it needs semantics and human intervention dependence constrained for accurate service discovery.

Automatic service discovery and selection can be prominent using semantic Web technologies [4],[5],[6]. The majority of current approaches to semantic web services for the Web service call to find that the information tagged through different approaches. However, this approach has several drawbacks. First, it is impractical to expect a number of new services to the descriptions of the material tagged. Second, most of the existing Web services using WSDL definitions and the associated objects are not included. Moreover, from the perspective of the service requester, the requester should be aware of all the knowledge that is in the domain. In particular, the service requester may be aware of all rules and regulations related to the service request. As a result, it should not be used for the service discovery process for many service requests.

In order to address the limitations of existing semantic based service discovery approaches, this paper presents an integrated self-adaptive semantic classification approach with web usage frequent patterns generated from web usage logs. Internet use data can be obtained from the server logs and can be analyzed to generate service information patterns to relate the semantic relation with the web domains for the various personalization and recommendation process[7][8]. This approach provides semantic annotation based on the service knowledge ontology which built by using Web services specific service knowledge and web usage frequent patterns for the accurate classification to improvise service discovery limitation.

The following paper organized as related work in Section-2, the proposed Self-adaptive Semantic Classification Approach in section-3. Section-4 presents the experiment evaluations and Section-5 presents the work conclusion.

Related Works
Discovery service automatically or semi-automatically by means of various IT services or service information in a particular context, which aims to restore the industrial informatics domain, is a growing research area. However, few studies into advertising information services planned industrial service advertising discovery using various heterogeneous and ambiguous features to find an advertising in this environment. The semantic web based web crawler is provide information using semantic web technology in an specific topics[12], [13]. Semantic technologies enhance the interoperability between heterogeneous elements which will
provide shared knowledge and technologies for widely used in industrial automation field [14],[15],[16]. Zheng et al. [17] proposed focused crawler based on a supervised ontology learning to maintain the harvest rates that are intended to maintain the crawler crawling process. The main idea of this is a Web crawler is to determine relatedness between ontology document and an artificial neural network (ANN) model. With a domain-specific ontology and an ontology of the title given to the representation of a concept and a set of concepts relevant to the topic ontology concept and other ideas are selected to represent the distance between the counter header background knowledge.

Web usage mining is the application of data mining methods for analyzing the usage of the domain website, utilizing web server logs. Web sites provides a wealth of useful information in form of Internet log usage for classification of Web service. This web service can support the efficient classification of improvisation is often used to create patterns [18],[19],[20]. Jespersen et al. [21] proposed visitor click-stream sequences based hybrid approach for analyzing the web usage. Hypertext probabilistic grammar on a partnership approach and the fact that the general line of the table can be used for tasks that used to mine in the blog. Mobasher et al. [22] proposed the web personalization system, the proposed mining-related tasks done offline usage data and automatic Web page unique online process on the basis of discovered knowledge. Chi et al. [23] proposed a analytical tool as LumberJack. It combine both user sessions and traditional K-means clustering algorithm for the statistical analysis through the use of user profiles[11]. Aabhas et al.[24] presents an approach for web service discovery based on semantic service categorization and semantic enhancement of the service request. It is based on the functional level of service classification ontology framework for achieving a solution. It is precisely utilize the function of the service on the basis of a set of Web services used for classifying. It includes the universal description for the discovery and integration for semantic-based categorization and classification in offline. Service request are better associated with the enhancement of relevant services semantically to achieves a good service matching. Service knowledge training approach for Web applications are based on service only, it does not adapt to the demand. In the proposed performs an object semantic classification using Web usage patterns of Internet service in the integration of knowledge and the need to resolve the classification approach.

Self-Adaptive Semantic Classification Approach
The process of web service discovery performed through efficient and accurate service web crawling in the process of classifying the Web services. A framework for self-adaptive semantic classification approach for service discovery is presented in Figure-1. It consists to two systems to perform the service discovery process as, Learning System and Classification System.

Learning System
Service Knowledge Learning
Service Knowledge Learning allowed to connect to the service metadata in parallel, the property offers a down-level concepts. Service metadata is generated, and store in a predefined service knowledge ontology for latter reference. The domain ontology of Web service share some common features of each concept is an abstraction to provide an array of ideas for this service offers. Supporting this notion, the essence of a concept and its sub-concepts ideas concepts are related. A high-level forum of all the metadata associated with the data by sub-concepts for the purposes of classification, the level of service concepts in one level, have the privilege of associating with the metadata service.

A service metadata module process the web service data collected from different domain from the web and those data utilized to builds a service knowledge ontology database, which will helps in the association classification to build log data patterns.

Web Usage Patterns Learning
Web usage learning based on the web usage mining, where the recorded web usage logs in Web server related to a user query is processed to generated service frequent patterns [25]. It undergoes data pre-processing to build frequent patterns using association rule mining process. Pre-processing is the removal of noise and irrelevant data, and in addition to this system is to provide a clean log data for association mining of Web service in combines with semantic information. Service frequent pattern are generated utilizing the pre-process data in a sequential association rule mining. In this case of association rule mining, the frequent patterns generated tend to maintain the sequence relation between the set of items discovered in relates the service knowledge ontology.

Classification System
The classification system executes two process to generate the classified web services from web data. In the initial process known as metadata information, will extracts the metadata information from the web data for the service classification.
The second process, executes the self-adaptive semantic classification process to relate the metadata information semantically with the service knowledge ontology and service frequent patterns association to find the accurate services discovered for user query request.

Service Knowledge Ontology Based Association

The self-adaptive classification approach implements the algorithm to relate the obtain meta information against the service knowledge ontology to associate semantically. Let’s assume that a service knowledge has a set of \( S = \{s_1, ..., s_k\} \) containing \( m \) Nodes containing \( t \) terms for semantically association and \( R_{rate} \) as relevancy rate as described in Algorithm-1.

**Algorithm-1: Service Knowledge Ontology Association**

\[
\text{Input:} \quad S \rightarrow \text{Service Knowledge Ontology data set} \\
M \rightarrow \text{Web Service Metadata} \\
\text{Output:} \quad R_{rate} \rightarrow \text{Web Services Metadata Classified Rate} \\
\text{Process:} \\
\begin{align*}
\text{For each set of } S & = s_1, ..., s_k \text{ do} \\
\text{Initialization: } & L_{count} = 0 \\
\text{For each term in } M & = t_1, ..., t_m \text{ do} \\
& \text{If } (M \text{ terms } c \text{ Nodes Objects}) \text{ Then} \\
& L_{count} = L_{count} + 1 \\
& \text{End If} \\
& \text{End For} \\
R_T = & L_{count} \\
R_{rate} = & \frac{R_T}{x \text{ terms}} \times 100 \\
\text{End For}
\end{align*}
\]

Based on the obtained \( R_{rate} \) obtained for each web service metadata are selected which are above the define relevancy rate threshold.

Service Frequent Pattern Based Association

The classified web service metadata is processed against the service frequent patterns to compute the best web service relevancy for a query as discussed in [26]. Let’s us consider that \( D_T \) represent the distinct number of terms in pattern \( T \). \( M_{total} \) represents the total number of distinct terms in metadata \( m \). \( N_T \) is the total number of terms in the patterns. \( T_{term} \) represents the total number of terms in patterns. \( T_{order} \) represent a pattern terms order in pattern, if terms order same as metadata then \( MT_{order} = 1 \) else it will be 0. Then, to compute the relevancy similarity of pattern equation (1) method will be used.

\[
\text{sim_relev}(T, m) = \frac{D_T + 1}{M_{total}} \times \frac{N_T}{T_{order}} \\
\]

The computed value of each web service metadata will be used for final ranking. Higher the value, higher will be the rank of the resulted web service.

**Experiment Evaluation**

To perform the experimental evaluation we create a database to evaluate the test using online web services published over internet and to create a service ontology structure as discussed below.

**Datasets**

Web servers and Web sites collect large amounts of data from the applications for the Web access and stored in log files. We collect a 5000 web log data from reachouthyderabad.com and customize the log according to our process requirement. A web log is a form of collection set of transactions of users and servers. Four well-defined domain services related to E-business, Healthcare, Travel and Industry ontology structure are defined. The defined structure will support to the automatic self-adaptive semantically classification and annotation of the web services with associated of ontology structure.

**Generated Service Frequent Patterns**

This process executes over the extracted log data and performs the sequential pattern generation in related to different user with different support threshold. The obtained results are shown in figure-2.

**Evaluation Results**

A 1000 different web pages we process offline to create metadata. This generated metadata used to build metadata ontology structure for semantic association in connection with the Internet service used with the system builds. The generated service metadata is stored in database for automatic semantic classification and categorization of the web services. Table-1 shows he metadata generated and there semantic classification relation and their categorization obtained during implementation. Table-2 shows the categorized web service links in related the service ontology structure.
Table 1: Classified WebPages based on Service Knowledge Ontology

<table>
<thead>
<tr>
<th>Webpage</th>
<th>Metadata Generated</th>
<th>Semantic Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page 1</td>
<td>100%</td>
<td>Category 1</td>
</tr>
<tr>
<td>Page 2</td>
<td>80%</td>
<td>Category 2</td>
</tr>
<tr>
<td>Page 3</td>
<td>70%</td>
<td>Category 3</td>
</tr>
<tr>
<td>Page 4</td>
<td>90%</td>
<td>Category 4</td>
</tr>
</tbody>
</table>

Table 2: Pattern Based Web Services Discovery Results

<table>
<thead>
<tr>
<th>E-Business</th>
<th><a href="http://www.gazono.com/">http://www.gazono.com/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry Equipment</td>
<td><a href="http://www.kenbros.com/">http://www.kenbros.com/</a></td>
</tr>
<tr>
<td>Security Equipment</td>
<td><a href="http://www.rimom.com/">http://www.rimom.com/</a></td>
</tr>
</tbody>
</table>

To evaluate the effectiveness of the proposed system we measure the Precision rate and Recall rate as defined [27]. The analysis is made on the 1000 WebPages obtained from web with different domain queries. With a variation support threshold association value from 10 to 100 we compute the measure in relate the number of relevant Data Semantically Matched, number of Relevant Data Not-Retrieved and number of Non-Relevant Data Retrieved. The computed measures are shown in figures-3 and 4. The obtained results show an improvisation in precision rate and minimize in the recall rates, which prove the improvisation of the proposal.

Figure 3: Precision rate at different threshold value

Figure 4: Recall rate at different threshold value

Conclusion

This paper presents a self-adaptive service discovery approach in a web environment. It defines an efficient approach to build a service ontology structure and classifying the web service into category for user service improvisation and frequent pattern mining sequence patterns, we investigated the effect of semantic information. The main functions of the proposed work is to semantically discover the service information from the Web pages by parsing, annotating, and storing their service information which will be used for classification the categorization of the web service based on specific service ontology domain knowledge. This approach defined a general format for service metadata and service concept, which enables the function of similarity computation and the association between metadata and concepts. The experiment results shows that, increase of the support threshold value can reduce the amount of associated and non relevant metadata, and the relatively higher support threshold values can benefit the overall performance of the automatic service discovery.

References


