Journal of Web Engineering, Vol. 14, No.3&4 (2015) 196-214
© Rinton Press

MACHINE LEARNING IN EFFICIENT AND EFFECTIVE WEB SERVICE DISCOVERY

KAILASH CHANDER BHARDWAJ

Department of Computer Science and Engineering
Thapar University, Patiala- 147004, India
Email: kailashbhar@gmail.com
Tata Consultancy Services

R.K. SHARMA

School of Mathematics and Computer Applications
Thapar University, Patiala-147004. India
Email: rksharma@thapar.edu

Received April 21, 2014
Revised October 29, 2014

The web service discovery mechanism has continuously evolved during the last years. There is plethora of information available about various techniques and methods used for meeting the challenge of improving web service discovery. A tremendous effort has been reported in literature and researchers are still contributing to make the web service discovery more effective and efficient. This paper discusses various eminent researchers' work in this direction using machine learning based techniques. Machine learning is a promising area for researchers to produce accurate estimates consistently. Machine learning system effectively "learns" how to estimate from training set of completed projects. We hope that this paper would benefit researchers to carry further work discussed in this paper and provide an outlook for the future research trends.

Key words: Semantics, web ontology language, web service description language, web services modeling language, web service modeling ontology, neural networks, fuzzy logic, ontology, quality of services.
Communicated by: B. White & C. Bizer

1 Introduction

Multiple web services integrate and interoperate to perform tasks, provide information, carry business activities and perform actions based on user's demands. The web service paradigm brings a number of advantages to application developers and end-users. The web service model simplifies business application development and interoperation, as it entails code reuse and loose coupling between services. It also enables users to choose, configure and assemble their own web services.

However, in order to employ full potential of web service, the web service paradigm must be supported by an efficient discovery mechanism. Web service discovery is the most important task in the web service model as it is of no use if a web service can not be located. Figure 1 shows the web service model with interaction between a service requester, service provider, and a service discovery system. The service providers are companies or organisations who offer web services with specialised functions or business operations. The web service registry is a broker that provides registry and search functions using a standard Universal Description, Discovery and Integration (UDDI) [50], which is the mostly used mechanism to find services on the web.

The Web service consumer searches a web service in UDDI registry and submits requirements using keywords. The keyword based techniques used in common search engines are not efficient as many irrelevant services may be included in the description of the queried keywords, leading to low precision. Also, the queried keywords may be semantically equivalent but syntactically different from the words in the offered services, leading to reduced recall. The key issue is that keywords are a poor way to capture the semantics of a service request or service advertisement. Thus a different mechanism is needed, one that entails locating web services on the basis of the capabilities they provide. .

One can search a service from UDDI using syntactically keyword-based search and category-based browsing of web services. UDDI is supported by major organizations and is already used in many tools. UDDI supports a generic framework for describing web services. There are various models for describing web services such as web service description language (WSDL) [51], web service modeling language (WSML) [53], web service modeling ontology (WSMO) [52] and web service modeling ontology for semantics (OWL-S [6]). WSDL, WSMO and OWL-S are based on the semantic web initiative. Semantic web services are seen as one of the most promising research directions to improve the integration of applications within and across enterprise boundaries. Both WSMO and OWL-S have the aim of providing the conceptual and technical means to realize semantic web services, improving the cost effectiveness, scalability and robustness of current solutions. WSMO provides a direct link between a web service, its capability and its interfaces, including the service choreographies and groundings. WSMO relies on four major elements namely, Ontologies, Goals, Web Services and Mediators. OWL-S is a service description framework that provides both rich expressive descriptions and well-defined semantics. OWL-S describes the characteristics of a service (upper ontology) by using three top-level concepts, namely, ServiceProfile, ServiceGrounding, and ServiceModel. OWL-S is also designed in such a way as to extend UDDI for service discovery.

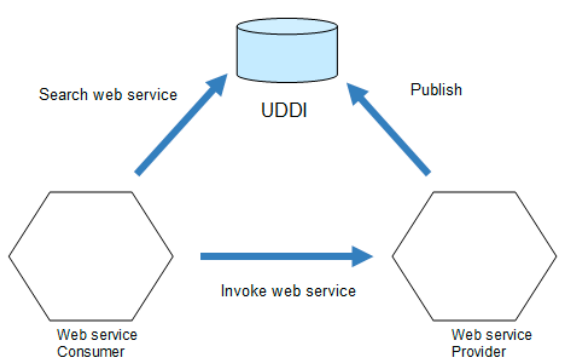

Figure 1 Web service discovery model

Recently, machine learning models have been playing a significant role in web service discovery mechanisms and researchers have shown that use of these models improves classification accuracy when applied on the service data. Web service features can be easily transformed in the format which can be used in machine learning algorithms. There are many proposed approaches in literature using machine learning methods for the web service classification area. But, still automated web service classification is a hot area of research and there is ample scope for improvement of efficiency and effectiveness with the proposed approaches. In this paper we will discuss various researchers work and their approaches for web service discovery using machine learning models.

This paper is structured as follows. In section 1, we briefly present the introduction about the web service discovery and various approaches to publish and discover web services. In section 2, we briefly describe current challenges in web discovery. Section 3 describes various machine learning techniques in web service discovery process. In Section 4, we discuss un-supervised machine learning techniques used by researchers and section 5 discuss about supervised machine learning techniques used in web service discovery. Section 6 presents comparison of performance of machine learning algorithms for web service and section 7 concludes this work and also contains future scope of work.

2 Web Service Discovery Current Trends and Challenges

Currently, web services are available in repositories such as UDDIs and Web portals (e.g. Xmethods [62], webservicex [63], webservicelist [64] etc). There exist three main approaches for the discovery of potential web services:

i) Using centralized repositories such as UDDI's
ii) Using web crawling techniques
iii) Search in specific web portals

Though there are not many public UDDI's available, but organizations are using option of discovery of web services through private UDDI Business Registries (UBR's) and search engines which are maintained by the individual organisations. There are various syntactic based standards such as WSDL, SOAP, and UDDI to support discovery of web services. Seekda (http://seekda.com) [68] is currently the most comprehensive global search engine for public web services.

2.1 Current Challenges in Web Service Discovery

Web services discovery is a challenging area and finding the desired service based on consumer's requirements is still a pain area. Some of the current challenges are:

- Quantum of services available on internet
- Non standardization of WSDL format
- UDDI matching supports only keyword matching that does not allow retrieval of Web services with similar functionality.
- Development of a common ontology for myriad of web services is difficult and cumbersome task.
- Manual annotation of web services.

2.2 Research Trends in Web Service Discovery

As discussed in this paper, due to the limitations of UDDI registry, it has not been adopted widely as expected during its origin and evolution. Standards such as UDDI, WSDL and SOAP have been created to help discovery of web services, but due to syntactic nature of these standards, web service discovery is not efficient.

Today, quantum of research is being done to explore various techniques for the efficient web service discovery. Researchers are working with the semantic web technologies and intelligent techniques to enhance the discovery mechanism. Especially, semantic web service discovery aims to discover the best matched web service, it mostly depends on the measurement of the similarity degrees between service request and service advertisement.

Current vision of researchers is semantic web and they have proposed and devised many algorithms, architectures and standards. Researchers have applied ontologies to web service discovery and found it is the most promising approach to semantically enrich web services and improve web service standards and bring structure to web services. The ontologies based standards like OWL-S and WSMO, DAML are used to describe web service and its resources. They replace existing web services by incorporating semantics which is understandable by machines.

Since there are many functionally similar web services available in the web, it is an absolute requirement to distinguish them using a set of non-functional criteria such as Quality of Service (QoS). So, current research also includes QoS based web service discovery. As is focus of this paper, machine learning techniques are widely used for web service discovery problem and researchers are using machine learning techniques on web service data WSDL, semantically annotated services to make discovery more efficient and effective.

3 Machine Learning Techniques in Web Services Discovery

In literature different architectures using machine learning have been proposed by some researchers to enhance web service discovery process. In web service discovery process, we find mainly three types of machine learning models widely used namely, supervised, unsupervised and semi - supervised methods as given in Figure 2. Subsequent sections provide more details on these machine learning models in web service discovery. The most popular models include Naïve Bayes (NB [54]), Decision Tree (DT) [5], Decision Rules (DR) [59], Association Rules (AR) [58], Neural Networks (NN) [56], and Support Vector Machines (SVM) [55].

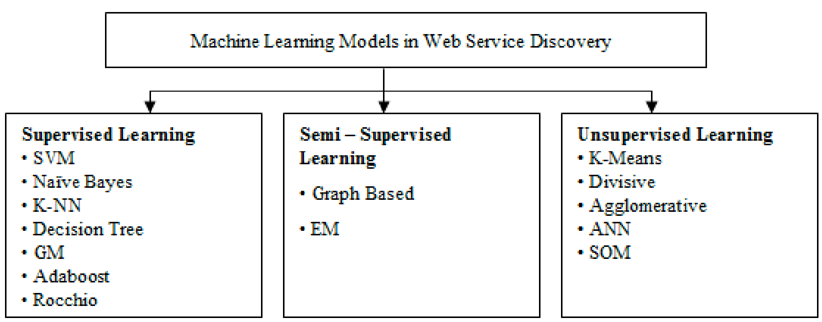

Figure 2 Machine learning models in web service discovery

Figure 3 shows the use of machine learning model in web service classification process. In the process of service classification a model is trained e.g. using the SVM method. A feature vector is created by extracting the functional descriptions, such as the input and output information, arguments etc. from the WSDL document of web services and used for the training data. In some scenarios non-functional features also called QoS are also extracted to improve the process. Extracted features are preprocessed using tokenization vector, stemming, filter stop words etc. to create the better training data. These created feature vector are fed to the machine learning model for classification.

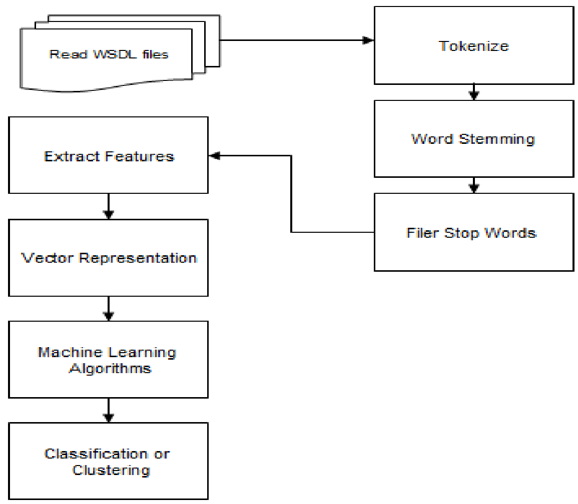

Figure 3 Machine learning model in web service classification process

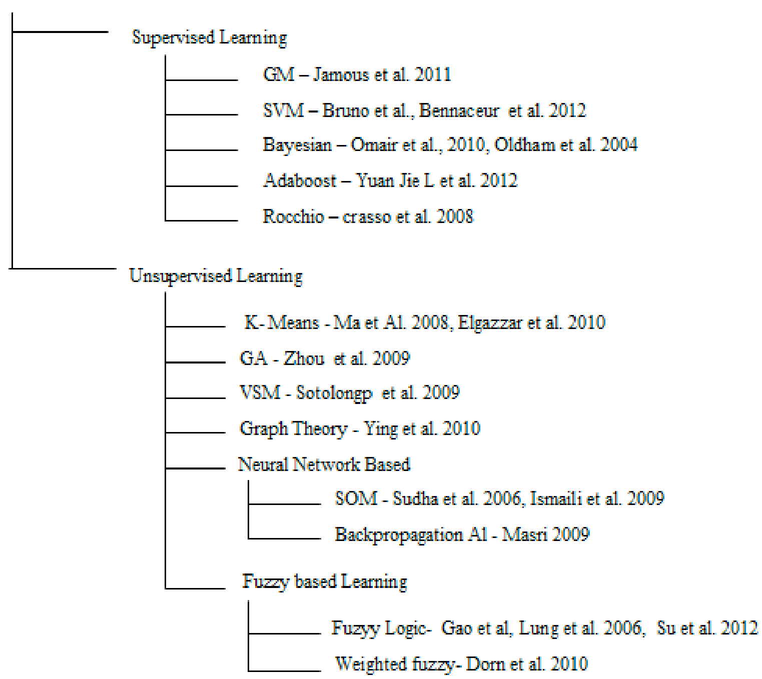

Figure 4 Taxonomy of web service discovery mechanisms using machine learning methods

Figure 4 shows the taxonomy of web service discovery mechanisms using machine learning methods. The proposed approaches can be classified based on the type of machine learning technique used for matchmaking. We have shown taxonomy for supervised and unsupervised based machine learning methods.

4 Unsupervised Machine Learning Techniques in Web Services

Unsupervised learning techniques are used to classify web services based on their functional and non functional similarities to enhance the web service discovery process. In unsupervised methods three categories, namely, non-hierarchical clustering, hierarchical clustering and agglomerative have been used for the discovery of web services.

The most widely used non-hierarchical clustering methods are *k*-Means and *k*-Medoids [7]. These methods require the user to specify the number of clusters, *k*, and prior to a cluster analysis. Hierarchical clustering is the most commonly used method for identifying closely related groups and uses distance matrices as clustering criteria. These methods can be agglomerative or divisive. The agglomerative methods proceed by a series of successive fusions of the individuals into groups, whereas divisive methods separate the individuals successively into finer groupings. Both complete linkage and single linkage are agglomerative methods. The single linkage, also called nearest neighbor, defines the *distance* between one cluster and another cluster as the shortest distance from any member of one cluster to any member of the other cluster. The complete linkage, also known as furthest neighbor defines the distance between one cluster and another as the longest distance from any member of one cluster to any member of the other cluster.

It is important to eliminate irrelevant services to improve the accuracy of the service matching process. Jiangang et al. [3] used novel clustering semantic algorithm *(k-means approach*) to eliminate irrelevant services. They utilized Probabilistic Latent Semantic Analysis (PLSA), a machine learning method, to capture the semantics hidden behind the words in a query, and the descriptions in the services, so that service matching can be carried out at the concept level.

Web service usage patterns and pattern discovery through service mining are used in [32]. They defined different levels of service usage data: i) user request level, ii) template level and iii) instance level. At each level, they investigated patterns of service usage data and the discovery of these patterns. These web service patterns, pattern discovery and pattern mining supported the discovery and composition of complex services for the increasingly complex business processes and applications.

Weighted bipartite graph is used in [12] to match user requests and published web service to enhance its discovery ability. It paid attention to the concept similarity of user requests. It avoided traditional keyword distance computing by using topology and reduced computing complexity and improved service discovery efficiency. It Introduced web service discovery method based on keyword clustering and concept expansion, mainly from the content of web service, reasoning of service request and service matching through classification.

Zheng and Bouguettaya [13] addressed the challenge of combinatorial explosion and developed scalable web service mining framework. This framework enabled the proactive discovery of interesting and useful service compositions. It discussed how interestingness and usefulness can be objectively evaluated. They presented a use case of a novel application of proposed framework to the discovery of pathways linking biological processes. Their future work included improving the agility of framework to accommodate the dynamic expansion and evolution of WSML services.

A novel technique to mine non semantic WSDL documents and cluster them into functionally similar web service groups has been proposed in [12]**.** It identified five key features that are extracted and integrated in order to group web services into functionality-based clusters. A search engine's crawler crawled WSDL documents from the internet and applied the proposed clustering approach offline in order to group similar functionally services. When a user queried the service search engine for a desirable objective, it used the clustered web services to match semantically the query and returned the most relevant web services that satisfied the requested objective. This approach has shown good performance for clustering web services compared with previous approaches. They proposed further to improve features integration by choosing optimized weights for each feature using a linear programming approach.

Two learning machines namely, Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA) have been used in information retrieval to learn latent factors from the corpus of service descriptions and group services according to their latent factors as given in [4]. They treated all the descriptions and attributes as text, semantic relations could be exploited to enhance the discovery and recommendation result based on the constructed clusters. They suggested that future work would focus on developing a query mechanism based on latent factors rather than matching key words in a query to the service descriptions' text. This approach could be applied to other service description models such as WSMO and this will also be investigated in future work. The probabilistic methods used (i.e. LDA and PLSA) could be trained using a small percentage of the whole dataset. The semantics of service descriptions would also be used for further enhancement of clustering results. The reasoning mechanisms would also be incorporated in the recommendation and discovery methods to provide more relevant results to service consumers.

Semantics is extensively used in recent works. This supports to overcome problem of keyword matching which is not able to provide desired matching web services and only works on the exact match of words. Semantics based representation of web services provides concept level matching of user requirements. Both OWL-S and WSDL are languages used to represent semantics descriptions of web services. A semantic web service mining model that allowed finding services from existing services dynamically using OWL-S ontologies has been presented in [18]. This work detected the services based on semantic relevance and this semantic relation was identified by the ontology analysis. An online search model has been proposed to detect the semantic relevant web services on the web. They mainly concentrated on the support of discovery in OWL-S IDE and changes to the existing UDDI registry. The techniques proposed in their work provided algorithms for the efficient use of OWL-S ontologies in UDDI, which could be easily applied to any OWL ontology. The intrinsic worth of using metrics based Models, WordNet metrics and semantic similarity metrics are explored in [20] to classify an ultimate approach for identifying web service resemblance with the help of metrics.

The semantic description of web services results into effective web discovery. Nayak and Lee [5] extended the semantic representation of services and grouped the similar web services in order to improve the service discovery using the semantic web services clustering (SWSC) method. SWSC leveraged OWL-S ontology and WordNet lexicon to enhance the description with semantics. SWSC utilised the clustering algorithm to group web services into domains for enhancing searching. The similarity functions not only considered the matching of text but also the context of the terms used. The context captured the semantic connections as similarity links between web services. In the system of Sotolongo et al. [29], an algorithm based on information retrieval that applied the lexical database WordNet together with a linear discriminant function has been proposed. This algorithm calculated the degree of similarity between words and their relative importance to support the development of distributed applications based on web services. This algorithm used the semantic information contained in the Web Service Description Language specifications and ranked web services based on their similarity.

As service discovery can be improved significantly using classifications, this approach is used in [6]. It classified the services by the graph theory clustering algorithm. I/O parameters of semantic web service have been represented by using Ontology. They have shown significantly improved efficiency of discovery as services discovery required not to compare all the services in the given services set.

QoS has gained attention of both researchers and academia and is successfully playing an important role in web services. Earlier, researchers focused more on functional aspects of web services. But recently, QoS is an important aspect of web services for both consumers and providers. The consumers want best selection of web service based on its described QoS. Zhou and Li [12] proposed a QoS-based service discovery framework along with entropy-based adaptive algorithm to realize service clustering. They presented an effective service discovery algorithm and the results of simulation experiments demonstrated the effectiveness of the proposed algorithm.

A service-on-demand discovery process is proposed in [23]. To meet the target, FCM clustering was adopted for agglomerative clustering between the user's QoS demand information and the QoS information of web service resource. The sequence could be determined by similarity computation in the same classification clustering. They presented hat the service-on-demand can be discovered efficiently to optimize network resources and improve the efficiency. A filtering process based on clustering techniques that narrowed the set of potential services has been introduced in [28]. The filtering process was based not only on the functional aspects of services, but also on non-functional ones. One of the weaknesses of the filtering process was lack of high efficiency and its complexity as processing ontological descriptions and reasoning on them was extremely time-consuming.

Proposed Approach	**Authors**
Service pattern mining	Qianhui et al. (2006) [32]
Semantics and clustering	Nayak and Lee (2007) [5]
Filtering and clustering	Abramowicz et al. (2007) [28]
Keyword clustering and ontology	Zhou and Li et al. (2008) [14]
WordNet and Linear Discriminant Functions	Sotolongo et al. (2008) [29]
Clustering semantic approach	Jiangang et al. (2008) [3]
Service clustering	Zhou and Li (2009) [12]
Service mining on web	Zheng and Bouguettaya (2009) [13]
Web service clustering using text mining techniques	Wei and Wilson (2009) [9]
Algorithm for semantic web services clustering and discovery	Ying (2010) [6]
Probabilistic methods for service clustering	Gilbert et al. (2010) [4]
Clustering WSDL documents	Elgazzar et al. (2010) [8]
Semantic web mining	Ramu et al. (2011) [18]
Similarities using matrices	Chandramohan et al. (2011) [20]
QoS oriented cluster algorithm	Feng et al. 2011 [23]

Table 1 Unsupervised machine learning based web service discovery

4.1 Neural Networks and Fuzzy Logic based applications in web service discovery

The Artificial Neural Networks (ANN) and Fuzzy logic have been playing a role in making the service discovery process efficient. An ANN is made up of artificial neurons that are connecting with each other. There have been various frameworks proposed based on fuzzy logic and fuzzy matching algorithms as summarised in this paper. A good number of articles have appeared recently for the efficient and effective selection of service using ANN. This section provides summarized view of literature for web service discovery based on ANN and Fuzzy Logic (refer- Table 2).

Proposed Approach	**Author**
A clustering based approach	Ram et al. (2006) [7]
Fuzzy web services discovery	Huang (2006) [26]
A novel approach	Ismaili et al. (2009) [10]
Neural network	Al-Masri et al. (2009) [11]
Intelligent clustering	Gao and Stucky (2009) [34]
Multithreaded fuzzy logic	Shehzad and Javed (2010) [16]
Weighted fuzzy clustering	Dorn and Dustdar et al. (2010) [24]
Ontology based fuzzy clustering	Gholamzadeh and Taghiyareh (2010) [33]
Fuzzy Logic and multi-phase matching	Su et al. (2012) [17]

Table 2 Neural networks and Fuzzy Logic based service discovery

Ram et al. [7] proposed an approach using unsupervised artificial neural network and empirically evaluated the approach using real web service descriptions. They have combined clustering techniques with string matching and leveraged the semantics of the XML-based service specification in WSDL documents. They have used q-grams string matching method for creating input features for cluster analysis. They further suggested other approaches using vector-space model etc. for string/document matching. They used a single clustering method for each experiment and further suggested that future research may evaluate the effects of combining multiple clustering methods. The proposed clustering based approach can be integrated with keyword search and predefined categorization based browsing, so that users can combine multiple search strategies in a flexible manner.

The existing discovery techniques do not take into account the diverse preferences and expectations of service consumers and providers which are generally used for searching or advertising web services. Huang et al. [26] presented a moderated fuzzy web service discovery approach to model subjective and fuzzy opinions, and to assist service consumers and providers in reaching a consensus. The method achieves a common consensus on the distinct opinions and expectations of service consumers and providers. This process is iterative such that further fuzzy opinions and preferences can be added to improve the precision of web service discovery.

A new approach that combines GHSOM (Growing Hierarchical Self Organizing Map) clustering technique with Latent Semantic Indexing is proposed in [10]. The overall process of discovering web services included: web services information processing, clustering heterogeneous services according to semantic similarity and ranking web services by their relevance. They suggested that further research should concentrate on the building tool for evaluation of the effectiveness of proposed method and to embed the SSDL in UDDI registries.

The backpropagation based neural network has also been proved to be useful for discovering web services of interest based on service classification as proposed in [10]. The use of neural networks provided a way to optimize the selection of the best available web service. The average performance rate of the neural network was 95%. They have used non-functional (QoS) properties of web services (response time, throughput, availability and compliance) to improve the probability of having relevant output results. Results suggested that the use of neural networks in discovering the most suitable web service is quite encouraging. It was observed that the backpropagation neural network has taken long time during the training mode due to the large data size which could become an issue when implementing such system in real-time manner. In addition, the ability of the system configuration to quickly adapt to current data being fed into it might become infeasible since the proposed method defined the number of hidden nodes prior in the training mode. They suggested the ability to use other types of neural networks such as ARTMAPs, Fuzzy ARTs, or Self Organized Map could be explored in the future and compared against the backpropagation algorithm.

The hybrid approaches based solutions are widely used in web service discovery as proposed in [34]. They populated a taxonomy structure by combining human knowledge with the technique of artificial intelligence. The web service was converted into standard vector format through WSDL document. The self learning neural network based learning algorithm and clustering algorithm were designed and implemented to classify the web services automatically.

Shehzad and Javed [16] proposed web services mining framework based on fuzzy set and rules satisfaction that proactively uncovers the interesting and useful individual web services and composes existing web services into composite web services. Fuzzy set has been generated based on the specified scope and weights have been assigned to each member of fuzzy set on basis of probability calculation in order to optimize the mining process. Proposed approach provided more efficient and precise results as long as system matures. The framework has been scalable with the growing number of web services repositories and provided efficient mining results. To mine huge web service repositories efficiently, multithreaded approach has been applied where a separate thread initiated for every member of the fuzzy set. This parallel processing approach for mining web services has improved the performance and made the framework scalable with growing web service search space.

A new fuzzy semantic clustering algorithm which assisted in discovering a group of similar web services through an individual query has been proposed in [33]. The ontology used supported the semantic analysis. Other approach which addressed the problem of limiting the number of required services that fulfill the required capabilities while exploiting the functional specialization of individual services is shown in [24]. This approach stroked a balance between finding one service that matches all required capabilities and having one service for each required capability. They have introduced a weighted fuzzy clustering algorithm that detected implicit service capability groups. The clustering algorithm considered capability importance and service fitness to support those capabilities. Evaluation based on a real world dataset successfully demonstrated the effectiveness and applicability for service aggregation.

A combination of multiple techniques are used for service discovery framework which incorporated semantic web, fuzzy logic, linguistic variable and multi-phase matching to get the most appropriate service of consumer's crisp request [17]. The main advantages of the proposed discovery method and matching mechanism have been: (i) it can formally describe not only the capability information, but the vague information of web service and implemented approximately reasoning based on ontology semantic, linguistic variable and fuzzy logic. (ii) multi-phase matching have been executed on different service abstract level, which could improve the efficiency and accuracy of service discovery.

5 Supervised Machine Learning Techniques in Web Services

The supervised algorithms use the training data, where each web service is labelled by categories. These algorithms analyze the training data and produce an inferred function, which can be used for classifying new data. This section is about the supervised classification techniques used in web service discovery. The automatic classification of web services into predefined categories has extensively been used by researchers. Many machine learning approaches such as Bayesian Classifier, Decision Tree, K-Nearest Neighbour (KNN), Support Vector Machines (SVMs), Neural Networks, Latent Semantic Analysis and Genetic Algorithms etc. are studied for classifying web services.

The automatic determining of service category is useful in semantics annotation and matching of web services. Crasso et al. [42] presented an AWSC (Automatic Web Service Classification) framework. They used relationship between category of web service and standard descriptions. They used text mining and machine learning techniques combination to improve classification precision.

The web service discovery mechanisms are using semantics based techniques extensively. Shafiq et al. [41] highlighted the fact that complex semantic descriptions to provide adequate information and reasoning methods take significant processing time. They came up with a hybrid approach with combination of semantics and machine leaning to improve the web service discovery process. They used non-functional properties of web services as light-weight semantics and used Bayesian classification technique to classify web services dynamically based on the light-weight semantics. They have shown that their approach is simple and efficient in terms of processing time.

The existing methods for automatic web service classification only consider the case where the category set is small. When the category set is big, the conventional classification methods usually require a large sample collection, which is hardly available in real world settings [35]. They introduced a new feature selection method to improve classification efficiency. This novel method classifies medium or big category set by utilizing the descriptive information of the sub-categories as the sample data. They presented a new feature selection based on semantic similarity of concepts to reduce feature space dimension, without the loss of the document features.

A good analysis and comparison of classification methods - Naïve Bayes, Markov blanket and Tabu, to search and rank web services has been done in [37]. They concluded that Naïve based Bayesian network performs better than other two techniques, Markov blanket and Tabu. Support Vector Machine (SVM) algorithm is used in [38] for the classification task. They proposed web service mining framework utilizing annotated capability specification as a key feature for classification, indexing and ranking. They used Semantics (OWL-S) service capability specification for automated and effective mining of web services. The ranking of web service was performed efficiently using semantic association of capability specification.

A generic non-functional based web services classification algorithm is used in [36]. This classification algorithm depended on information supplied by web service provider at the registration time. They proved mathematically and experimentally the usefulness and efficiency of proposed algorithm. Semantic annotation of services plays an important role for service classification. The challenge is annotation should be in agreement to a specific ontology and service description should be related too other services descriptions [44]. They proposed an approach to automatically classify services to specific domains, identify key concepts inside service descriptions and build a lattice of relationships between service annotations. Results obtained are promising for getting useful insights in service publication and retrieval.

The service interface descriptions are used in [45] for automatic classification of web service and this can speed up the service matching procedure considerably. They used machine learning based on

SVM to build categorizers of web service interface descriptions. They evaluated number of machine learning methods and SVM gave the best results and how the performance is influenced by the design of the feature extraction component.

The researchers have realised that more expressive descriptions of web services is the need of hour. Most of recent approaches suggest that ontologies should be used to describe web services and annotation purpose. Researchers felt need of automatic annotation mechanism and [46] created a framework call METEOR-S. They used Naïve Bayesian classifier and showed that their framework can match ontology faster and improved accuracy.

A new approach of grouping web services into functionally similar clusters by mining web service documents and generating an ontology via hidden semantic patterns present has been proposed by [47]. It proposed an approach to identify the cluster center that combines service similarity with the term frequency–inverse document frequency values of service names to improve utility of clusters. Experimental results show that proposed clustering approach performs better.

The ensemble methods have also been used widely for web service discovery problem. These algorithms are learning algorithms that combines a set of classifiers and then classify data by taking a weighted vote of each classifier predictions. A new approach is proposed in [40] which apply automatic web service semantic annotation. They evaluated the accuracy of the Ensemble Learning Classification method and find its accuracy much higher than any other classification method. It is shown that annotation and the mapping between concepts and ontology directly affected the result of the classification accuracy and after manual calibration the accuracy of each classifier was improved to some degree. They emphasised that WSDL file annotation is necessary to improve accuracy and efficiency of classification which enables better accuracy of web service discovery.

The AdaBoost ensemble algorithm which is often referred to as the best out-of-the-box classifier has also been used for the web service classification. AdaBoost is short form of “Adaptive Boosting" and is a machine learning meta-algorithm and can be used in conjunction with other learning algorithms to improve performance. Varguez-Moo et al. [60] proposed a web service discovery approach using machine learning algorithms Naïve Bayes, SVM and Adaboost to classify web services. They used QoS parameters into account as semantic information. They concluded that AdaBoost algorithm had the best mean percentage of precision and Naïve Bayes algorithm provided best mean percentage of recall and accuracy.

Saha et al. [67] used Tensor Space Model and Rough Ensemble classifier for web service classification. They have used two steps for improvement for web services classification results. In the first step they used tensor space model and achieved better classification results over existing results. In the second step they further improved classification results by using Rough set based ensemble classifier.

Proposed Approach	**Author**
Non functional classification of web services	Jamous et al. (2011) [36]
Classification based on Bayesian network	Mohanty et al. (2012) [37]
Annotated Capability Specifications	Satish and Wahidabanu (2012) [38]
SVM and concept based classification	Bruno and Canfora (2005) [44]
SVM based classification	Wang et al. (2010) [35]

Ensemble classification	Yuan-jie and Jian (2012) [40]
Light weight semantics and Bayesian classification	Shafiq et al. (2010) [41]
Rocchio classifier	Crasso et al. (2008) [42]
Web Service Interface based classification	Bennaceur et al. (2011) [45]
Web Service Interface based classification	Oldham et al. (2004) [46]
Hybrid Ontology Learning	Banage et al. (2013) [47]
AdaBoost Ensemble approach	Varguez-Moo et al. [60]
Roug Ensemble approach	Saha et al. [67]

Table 3 Supervised Machine Learning based service discovery

Contributors	***Approach***	***Semantic support***	***QoS support (y/n)***	***Ontology language***	***Ranking method (y/n)***	***Machine learning method***	***Annotation capability***	***Self learning***	***NLP used***
Lu (2005) [15]	Unsupervised	√	×	×	√	×	×	×	×
Ram et al. 2006 [7]	Unsupervised	√	×	×	×	NN	×	√	×
Nayak and Lee 2007 [5]	Unsupervised	√	×	OWL-S	×	HA	×	×	×
Zhou and Lee 2009 [12]	Unsupervised	√	√	WSMO	√	GA	×	×	×
Sotolongo et al. 2008 [29]	Unsupervised	√	×	×	√	VSM	×	×	×
Ying 2010 [6]	Unsupervised	√	√	√	×	GT	×	×	×
Jamous et al. 2011 [36]	Supervised	×	√	×	×	GM	×	×	×
Su et al. 2012 [17]	Unsupervised	√	×	OWL	×	×	×	×	×
Mohanty et al. 2012 [37]	Supervised	×	√	×	√	NB, MB and TB	×	×	×
Satish and Wahidabanu et al. 2012 [38]	Supervised	√	×	OWL-S	√	SVM	×	×	×
Wang and Stroulia 2010 [35]	Supervised	√	×	×	×	SVM	×	×	×
Shafiq et al. 010 [41]	Supervised	√	√	×	×	Bayesian	×	×	×
Yuan-jie and Jian 2010 [40]	Supervised	√	×	√	×	Ensemble learning-Adaboost	√	×	√
Bruno and Canfor 2005 [44]	Supervised	√	×	√	×	SVM	√	×	×
Bennaceur et al. 2012 [45]	Supervised	×	×	√	×	SVM	×	×	×
Oldham et al. 2004 [46]	Supervised	√	×	√	×	NB	√	×	×

Table 4 Comparison of machine learning approaches for web service discovery with various parameters

Naïve Bayes - NB
Markov Blanket – MB
Tabu Based - TB
Support Vector Machine – SVM
Generic method – GM
Entropy-Based – EB

Hierarchical Agglomerative - HA
Neural network - NN
Graph theory - GT
Vector Space Model – VSM
Genetic Algorithm - GA
Natural Language Processing – NLP

The comparison of existing prominent machine learning based web service discovery approaches are shown in Table 4. These approaches are categorized based on parameters like Approach, Semantic support, QoS support, Ontologies, Ranking method, Machine learning method, Annotation capability, Self learning and NLP. It is evident that different approaches have been used by researchers while using combinations of different mechanisms. Analysis of the matrix suggests that most of the proposed approaches used various classification techniques along with semantics and ontologies to get better discovery results.

6 Performance comparison of machine learning algorithms in web service classification

Reference	Algorithm Used	Accuracy achieved	Remarks
Swami et al. [65]	Augmented Naïve Bayes	80.72 %	Augmented Naïve Bayes, Tree Naïve Bays, Sons & Spouses approach predicted better classification accuracy of 80.72% than other techniques.
	Naïve Bayes	79.89%	
	Tree Augmented Naïve Bayes	80.72%	
	Markov Blanket	72.73%	
	Augmented Markov Blanket	72.73%	
	Sons & Spouses	80.72%	
Bennaceur et al. [45]	K-NN	39.59%	They used text mining and machine learning for classifying web services. Results shows that Rocchio algorithms provides better accuracy than other algorithms.
	Naïve Bayes	79.38%	
	Rocchio	85.08%	
Heß et al. [66]	Naïve Bayes	86%	They find that an ensemble approach that treats web services as structured objects is more accurate than an unstructured approach.
	SVM	73%	
Yuan-jie et al. [40]	Naïve Bayes	69.71 %	Study shows that accuracy of the ensemble learning classification method is much higher.
	SVM	77.31 %	
	AdaBoost	89.15 %	
Mohanty et al. [37]	Naïve Bayes	85.62%	Study shows that Naïve based Bayesian network performs better than other two techniques comparable to the classification done.
	Markov blanket	81.36%	
	Tabu	82.45%	

Table 5 Performance of machine learning algorithms in web service classification

It is found that most of the web service discovery research focuses on the achievable accuracy of various machine learning algorithms used for the classification purpose. The different studies have shown that an accuracy of high value has been achieved by number of different algorithms. Table 5 shows the comparison of machine learning algorithms performances based on classification accuracy as observed from various studies.

7 Conclusion and Discussion

This paper provides a survey of machine learning models in web service discovery approaches. One can observe that worldwide tremendous development and exploration work has been done in the area of improving the web service discovery. It is a challenging task to effectively discover the desired services that logically or contextually matches the needs of consumers.

As evident from literature, current web service discovery mechanisms focus on how web services can be accessed rather than what are their offered capabilities or functions. Semantics can help to express capabilities of web services and can improve the searching for consumer for the exact match of their requirement. This match is further improved by narrowing down the web service space using clustering methods which should provide significant opportunities to make the search efficient.

In this survey paper, we have highlighted and summarized the key points of work done by various eminent researchers using conventional machine learning techniques. Most of the work presented suggested that machine learning methods along with keyword search and predefined categorization based browsing, Natural Language Processing techniques, semantics representation of concepts, searching based on non functional parameters can be implemented in different combinations to provide multiple search strategies in a flexible manner. We hope the study provided in this article will be useful to web service publishers to select an appropriate web service discovery mechanism. It should also help consumers to get their best match for the service request.

We conclude that despite the quantum of work done, much needs to be done for improvement of performance and accuracy of web service discovery match process. Our study encourages that no one machine learning technique can be considered as the best technique and we see scope of improvement in the existing researches and tools. We see a trend of hybrid approaches to address the web service discovery pain area. It is observed that following areas can be further worked upon as highlighted below:

- Web service feature selection methods can be improved for better web service classification process.
- Use of optimized classifiers to improve performance in terms of accuracy, recall, precision and time to train and testing.
- Better use of semantics and ontology for the web service classification, clustering and informational retrieval.
- Ranking methods to arrange selected web services per user’s requirement.
- Efficient and effective combination (hybrid approaches) of machine learning algorithms for better results.

Acknowledgement

The authors take this opportunity to thank the anonyms reviewers of the paper for the critical reviews and suggestions. This has helped us in improving the contents of this paper.

References

1. Paolucci, M., Kawamura, T., Payne, T. R., Sycara, K.: Importing the semantic web in UDDI. In proceedings of the International Workshop on Web Services, E-Business, and the Semantic Web. Toronto, May 27-28, pp. 225-236, 2002.
2. Wang, Y., Stroulia, E.: Flexible interface matching for web service discovery. In proceedings of the Fourth International Conference on Service Oriented Computing (WISE'03), Italy, IEEE Computer Society Press, pp. 147-156, 2003.
3. Jiangang, M., Yanchun, Z., Jing, H.: Efficiently finding web services using a clustering semantic approach. In proceedings of the International workshop on context enabled source and service selection, integration and adaptation, 2008.
4. Gilbert, C., Payam, B., Klaus, M.: Probabilistic methods for service clustering. In proceedings of International Semantic Web Conference (ISWC), 2010.
5. Nayak, R., Lee, B.: Web service discovery with additional semantics and clustering. In proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, pp. 555-558, 2007.
6. Ying, L.: Algorithm for semantic web services clustering and discovery. In proceedings of International Conference on Communications and Mobile Computing, 2010.
7. Ram, S., Hwang, Y. and Zhao H.: A clustering based approach for facilitating semantic web service discovery. In proceedings of 15th Annual Workshop on Information Technologies & Systems (WITS) Paper, 2006.
8. Elgazzar, K., and Hassan, A. E., Martin P.: Clustering WSDL documents to bootstrap the discovery of web services. In proceedings of the 2010 IEEE International Conference on Web Services, pp. 147-154, 2010.
9. Wei, L., Wilson W.: Web service clustering using text mining techniques. International Journal of Agent Oriented Software Engineering, vol. 3, no. 1, pp. 6-26, 2009.
10. Ismaili, F., Zenuni, X., Raufi, B.: A novel approach for efficiently finding web services on the web. In proceedings of the ITI 31st International Conference on Information Technology Interfaces, 2009.
11. Al-Masri, E., Mahmoud, Q. H.: Discovering the best web service: A neural network based solution. In proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2009.
12. Zhou, J., Li, S.: Semantic web service discovery approach using service clustering. In proceedings of International Conference on Information Engineering and Computer Science, pp. 1-5, 2009.
13. Zheng, G., Bouguettaya, A.: Service mining on the web. IEEE transactions on services computing, vol. 2, no. 1, 2009.
14. Zhou, J., Zhang, T., Hui, M., Xiao, L., Chen, G., Li, D.: Web service discovery based on keyword clustering and ontology. In proceedings of IEEE International Conference on Granular Computing, vol. 1, pp. 844-848, 2008.
15. Lu, H.: Semantic web services discovery and ranking. In proceedings of the IEEE/WIC/ACM international conference on web intelligence, 2005.
16. Shehzad, K., Javed, M. Y.: Multithreaded fuzzy logic based web services mining framework. European Journal of Scientific Research, vol. 41, no. 4, pp. 632-644, 2010.

17. Su, Z., Chen, H., Zhu, L., Zeng, Y.: Framework of semantic web service discovery based on fuzzy logic and multi-phase matching", Journal of Information & Computational Science vol. 9 , no. 1, pp. 203- 214, 2012.
18. Ramu, G., Dr Reddy, B. E.: A framework for semantic web mining model. International Journal of Internet Computing, vol. 1, no. 1, pp. 94-98, 2011.
19. Huang, A.: Similarity measures for text document clustering. In proceedings of the New Zealand Computer Science Research Student Conference, pp. 49-56, 2008.
20. Chandramohan, D., Khapre, S., Ashokkumar, S.: A study of finding similarities in web service using metrics. International Journal of Scientific & Engineering Research, vol. 2, no. 6, 2011.
21. Murali, K. S., Durga, B. S.: An efficient approach for text clustering based on frequent item sets. European Journal of Scientific Research ISSN 1450-216X, vol. 42, no. 3, pp. 399-410, 2010.
22. Adala, A., Tabbane, N., Tabbane, S.: A framework for automatic web service discovery based on semantics and NLP techniques. Advanced in multimedia-special issue on web services in multimedia communication, vol. 2011, no. 1, 2011.
23. Feng, X., Wua, G.-F.: A cluster algorithm for QOS-oriented supply and demand discovery. In proceedings of SIGSVC Workshop, Sprouts:Working Papers on Information Systems, 2011.
24. Dorn, C., Dustdar, S.: Weighted fuzzy clustering for capability-driven service aggregation. IEEE International Conference on Service-Oriented Computing and Applications, pp. 1-8, 2010.
25. Sharma, S., and Batra, S.: Applying association rules for web services categorization. In proceedings of International Journal of Computer and Electrical Engineering, vol. 2, no. 3, pp. 465-468, 2010.
26. Huang, C.-L., Lo, C.-C., Chao, K.-M., Younas, M.: Reaching consensus: A moderated fuzzy web services discovery method. Information & Software Technology vol. 48, no. 6, pp. 410-423, 2006.
27. Nayak, R., Tong, C.: Applications of data mining in web services. In proceedings of the 5th International Conferences on Web Information Systems, Brisbane, Australia, pp. 199-205, 2004.
28. Abramowicz, W., Haniewicz, K., Kaczmarek, M., Zyskowski, D.: Architecture for web services filtering and clustering. In proceedings of International Conference on Internet and Web Applications and Services, 2007.
29. Sotolongo, R., Kobashikawa, C., Dong, F., Hirota, K.: Algorithm for web service discovery based on information retrieval using WordNet and linear discriminant functions. Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 12, pp-182-183, 2008.
30. Web Services Architecture, http://www.w3.org/TR/ws-arch.
31. Makripoulias, Y., Makris, C., Panagis, Y., Sakkopoulos, E., Adamopoulou, P., and Tsakalidis, A.: Web service discovery based on quality of service, In proceedings of IEEE International Conference on Computer Systems and Applications, pp. 196-199, 2006.
32. Qianhui, L., Jen-Yao, C., Steven, M., Steven, O.: Service pattern discovery of web service mining in web service registry-repository. In proceedings of IEEE International Conference on e-Business Engineering, 2006.
33. Gholamzadeh, N., and Taghiyareh: Ontology based fuzzy web services clustering. In proceedings of International Symposium on Telecommunications, pp. 721-725, 2010.
34. Gao, H., Stucky, W.: Web services based intelligent clustering techniques. In proceedings of IEEE international forum on information technology and applications, pp. 242- 245, 2009.
35. Wang, H., Shi, Y., Zhouy, X., Zhou, Q.: Web service classification using support vector machine. In proceedings of IEEE 22nd International Conference on Tools with Artificial Intelligence, 2010.
36. Jamous, M. M., Safaai, B. D.: Web services non-functional classification to enhance discovery speed. International Journal of Computer Science Issues, vol. 8, no. 4, 2011.
37. Mohanty, R., Ravi, V., Patra, M. R.: Classification of Web Services Using Bayesian Network. Journal of Software Engineering and Applications, vol. 5, no. 4, pp. 291-296, 2012.

38. Satish, N., Wahidabanu, R. S. D.: Web service mining based on annotated capability specifications. European Journal of Scientific Research, vol. 69, no.3, pp. 416-427, 2012.
39. Paik, I., Fujikawa, E.: Web service matchmaking using web search engine and machine learning. International Journal of Web Engineering vol. 1, no. 1, pp. 1-5, 2012.
40. Yuan-jie, L., Jian, C.: Web service classification based on automatic semantic annotation and ensemble learning. In proceeding of IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, pp. 2274-79, 2012.
41. Shafiq, O., Alhajj, R., Rokne, J. G., Toma, I.: Light-weight semantics and bayesian classification: A hybrid technique for dynamic web service discovery. In proceeding of IEEE International conference on Information Reuse and Integration, pp. 121-125, 2010.
42. Crasso, M., Zunino, A., Campo, M.:AWSC: An approach to web services classification based on machine learning techniques. CONICET, Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. no. 37, pp. 25-36, 2008.
43. Kuang, L., wu, J., Deng, S.: Service classification using adaptive back-propagation neural network and semantic similarity. In proceedings of the 10th International Conference on Computer Supported Cooperative Work in Design, 2006.
44. Bruno, M., Canfora, G.: Using SVM and concept analysis to support web service classification and annotation. In proceedings of IEEE international conference on e-Technology, e- Commerce and e-Service, 2005.
45. Bennaceur, A., Issarny, V., Johansson, R., Moschitti, A., Sykes, D.: MachinelLearning for automatic classification of web service interface descriptions. In ISoLA 2011 Workshops, pp. 220–231, 2012.
46. Oldham, N., Thomas, C., Sheth, A., Verma, K.: METEOR-S web service annotation framework with machine learning classification. In proceedings of the 13th international conference on World Wide Web, pp. 553-562, 2004.
47. Banage, T., Kumara, G. S., Incheon, P., Wuhui, C.: Web-service clustering with a hybrid of ontology learning and information-retrieval-based term similarity. IEEE 20th International Conference on Web Services, 2013.
48. Erl, T.: Service-Oriented Architecture: A field guide to integrating XML and web services, Prentice Hall, 2004.
49. Josuttis, N.M.: SOA in Practice: The art of distributed system design. Publisher: O'Reilly Media, 2007.
50. Bellwood, T., Clement, L., Riegen, C.V., Uddi version 3.0.1. http://uddi.org/pubs/uddi v3.htm.
51. World Wide Web Consortium (W3C): Web Services Description Language (WSDL)1.1. http://www.w3.org/TR/wsdl/ 2001.
52. Fensel, D., Lausen, H., Polleres, A., Bruijn, J., Stollberg, M., Roman, D., and Domingue, J.: Enabling semantic web services: the web service modeling ontology, Springer, 2007.
53. Bruijn, J. d.: The web service modeling language WSML, WSMO final draft v0.2, 2005.
54. Sahami, M., Dumais, S., Heckerman, D., Horvitz, E.: A Bayesian approach to filtering junk e-mail, learning for text categorization, pp. 55-62, AAAI Workshop, 1998.
55. Vapnik, V. N., Druck, H., Wu, D.: Support vector machines for spam categorization, IEEE Trans. on Neural Networks, 10(5), pp. 1048-1054, 1999.
56. Wang, Y.Q.: Theory and method of artificial intelligence. Press of Xi'an JiaoTong University, 2001.
57. Quinlan, J. R.: Induction of Decision Trees. Machine Learning 1: 81-106, Kluwer Academic Publishers, 1986.

58. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD, 1993.
59. Pudil, P., Novovicova, J., Kittler, J.: Automatic machine learning of decision rules for classification problems. In proceedings of 4th BMVC93, 1993.
60. Varguez-Moo M., Moo-Mena, F., and Uc-Cetina, V.: Use of classification algorithms for semantic web services discovery. Journal of Computers, vol. 8, no. 7, pp. 1810- 1814, 2013.
61. Sonawani, S., and Mukhopadhyay., D.: A decision tree approach to classify web services using quality parameters, In proceedings of International Conference on Web Engineering and Application, ICWA, 2013.
62. XMethods, http://www.xmethods.net, accessed February 2012.
63. http://www.remotemethods.com, Accessed February 2012, Remote Methods: Home of Web Services.
64. Web Service List, http://www.webservicelist.com , Accessed February 2012.
65. Swami Das M., Mohanty, R., Vijayalakshmi, D., Govardhan, A. : Application of data mining using Bayesian belief network to classify quality of web services. Special Issue of International Journal of Computer Science & Informatics (IJCSI), ISSN (PRINT) : 2231–5292, vol. II, no. 1, 2008.
66. Heß, A., Kushmerick, N.: Machine learning for annotating semantic web services. In proceedings of AAAI Spring Symposium on Semantic Web Services, 2004.
67. Saha, S., Murthy, C.A., Pal, S.K.: Classification of web services using tensor space model and rough ensemble classifier. In Proceedings of 17th international Symposium on Foundations of Intelligent Systems (ISMIS'08), pp. 508–513. Toronto, Canada, 2008.
68. Kritikos, K., Plexousakis, D.: Semantic QoS-based web service discovery algorithms. In Proceedings of the 5th IEEE European Conference on Web Services, ECOWS 07, 2007.