

# Classifying Web Services based on QoS Parameters using Extended Dataset

Vandan Tewari  
S.G.S.I.T.S  
23, Park Road  
Indore

Urjita Thakar  
S.G.S.I.T.S  
23, Park Road  
Indore

Nirmal Dagdee  
Comp.Engg.Dept.,S.D. Bansal  
College of Tech. ,Umaria  
Indore

## ABSTRACT

Web Services are the interoperable and loosely coupled components for simplifying business processes over the web and are popular model of distributed computing now a days. There are a very few datasets available regarding Quality of Service (QoS) for web services since web services datasets are difficult to create. The reason being that it requires more effort to collect service information from various sources and structure it into a dataset. QWS dataset [5] is a data set which contains quality of service parameters for Web Services. However, this dataset includes only nine quality of service parameters and misses on some very important non-functional attributes such as security, interoperability and robustness which may be important while discovering web services which are sensitive in nature. This constitutes the class of services used for defense applications related to national security and service related to heavy financial transactions. In this paper, a method of web service classification has been proposed based on QoS Parameters of web services. Additionally an extended QoS parameters data set has been created using statistical techniques and concept of Highly Normalized Function. This extended dataset consists of unlabeled samples, which are processed to generate labeled dataset. This synthetic dataset has been named as Non-functional Parameters Dataset (NfPD).

## General Terms

QoS , Web Services

## Keywords

QoS ,Web Services, Dataset

## 1. INTRODUCTION

Web services[1] are the self-contained, XML [2] based components of an application which can be accessed remotely. A Web Service is defined by the W3C [3] as "a software system designed to support interoperable machine-to-machine interaction over a network". Web services can be accessed over a network, such as the Internet, and executed on a remote system hosting the requested services. They provide an interoperable and loose coupling of services. It solves the complex problems by allowing communication among various programs without user interfering. Web Services are the latest model of internet computing and are gaining wide popularity for business to business interactions over the web.

Web services have three main entities: Simple Object Access Protocol [4] (SOAP), Universal Description, Discovery and Integration (UDDI)[5] and Web Service Description

Language. SOAP is a lightweight XML based protocol for exchange of information in a decentralized, distributed environment. It is platform and language independent protocol. Using SOAP over HTTP allows for easier communication through proxies and firewalls than previous remote execution technology. UDDI is a registry for publishing services. It contains details of services and work as a directory structure. UDDI[5] is a platform-independent framework for describing services, discovering businesses, and integrating business services by using the Internet. WSDL is also XML based language which provides the mechanism through which Web services definitions are exposed to the world and to which Web services implementers need to conform when sending SOAP messages.

As web service become more popular, there are many factors like performance, reliability, security, availability, robustness, interoperability, scalability and other functional and non-functional factors which are emphasized more. Performance is the measure of response. Reliability is the ability of the system to keep operating. Security is a measure of system's ability to resist unauthorized attempts at usage and denial of service while still providing its services to legitimate users. Availability is the proportion of time the service is available for usage. It is measured by the length of time between failures as well as how quickly the system is able to resume operation in the event of failure. Robustness is the resilience of the service under stress or when confronted (to face difficult situation) with invalid input. It is the ability of the software system to maintain function even with the changes in internal structure or external environment. Interoperability is the capability of a service to execute among various functional units with only small or none changes. Scalability is the property of the service which indicates its ability to either handle growing amounts of work in a graceful manner. Service should not degrade because of increasing amount of work. Functional factors involve that whatever system has functionality it must not fail to provide. These qualities are directly impact on web services. While discovering and using the web services a user can benefit more if he is aware of its QoS properties. A QoS dataset of non-functional properties of web services can therefore be used in conjunction with classification and clustering for offering better services to a user. Data mining [6] is the extraction of required pattern from the large databases and the QoS dataset for web services may be useful for web service discovery used in conjunction with data mining. Generating QoS dataset from services is a cumbersome task because services should be discovered and their QoS behavior is to be observed over the time to calculate values of non-functional attributes. Dataset [7] is a collection of data which is basically stored in a tabular form. The data which is stored in the dataset can be of two types: labeled data

and unlabeled data. The data which belongs to a particular class is called labeled data. In other words, labeled data is the data which has features that are sufficient to help the classification of that data. Labeled data is difficult to collect because it is expensive, time consuming and require more human effort to be labeled. While Unlabeled Data does not have class labels and it requires less human efforts. [8] describes QWS dataset which is a collection of nine Quality of Web services parameters. This dataset is a labeled dataset. In this proposal additional QoS Parameters have been generated in addition to the nine parameters present in QWS dataset to form the Non-functional Parameters Dataset. These parameters are Interoperability, Security, Scalability and Robustness.

The remaining of this paper is organized as follows. Section 2 outlines the related work of creating dataset. In section 3, we present our method in order to create NfPD dataset and describe the overall system. Section 4 presents the results. Finally, in Section 5, conclusions and discussions have been presented.

## 2. RELATED WORK

The Web is the most valuable source of information in current scenario of information explosion. The organizations are setting their business on the Internet for their commercial, economical, educational and political growth. To support this growth, additional tools and supporting technologies are required that can help the development of these applications. So new concepts and technologies like Service Oriented Architecture (SOA) [9] is redefining the path in which the internet is being used. The communication in SOA can involve either simple data passing or it could involve two or more services coordinating some activity. Traditional distributed processing model faces the problem of non-interoperability like CORBA, EJB and RMI. Service Oriented Architecture (SOA) resolves it in effectively and efficiently. The concept of SOA is based on web services. Web services are the application component which is accessed remotely to provide distributed environment.

If similar services are presented to a user, it is difficult for a user to choose an appropriate service from them. Associating QoS with services can help us in offering better services to a client. Any business process performs well if related services perform well. The higher QoS ranking of the web services makes a service more popular with a client and hence if the ranking of the web services improves, business also benefits from it. The quality of the service can be measured by measuring the non-functional attributes of the service. The services which are designed for simple tasks and fulfills the functional requirements of a client are assumed to be successful services. However when complex services are designed, simply fulfilling functional requirements do not suffice. Functional requirements are used to judge the specific operations that service can perform while non-functional requirements are aimed at judging the behavior of a service in different usage environments. When the client requirements are complex, the issue of additionally observing non-functional requirements comes into picture. These non-functional attributes can actually decide the ranking of the web services. The services and their corresponding QoS values can be used applying classification and clustering of web services which can benefit an end user in service discovery [10]. In [11], web services have been classified

based on their QoS values to improve a client oriented web service discovery.

Data mining is the extraction of useful pattern from huge data collections. The web service QoS dataset bridges the rift between web services and data mining. The collection of qualitative attributes forms a dataset. In [7] a dataset has been described as a collection of labeled and unlabeled samples arranged in usually tabular form. Labeled data are the data whose classes are known and vice versa for unlabeled samples. There is a scarcity of datasets for research in web services and only a few datasets are available due to the reason that it requires more effort to collect services from the real world. Quality of Web Service (QWS) [8] dataset is a web services based dataset. In [12] and [13], authors have described QWS Dataset which rank the web services on the basis of their non-functional attributes. While QWS dataset consists of nine quality of service parameters which evaluates the quality of the web services, it has left some important parameters like Interoperability, Security, Scalability and Robustness. In this proposal these parameters have been synthetically generated to create the proposed synthetic dataset to be used for classification of web services to augment the web service discovery for a client. These parameters have been used for service ranking in this proposal using the web service relevancy function defined in [14]. On the basis of this WsRF ranking, the QWS Dataset is classified into four categories which are Platinum, Golden, Silver and Bronze however in this proposal the classification of web services has been done based on assumptions taken in [13] by the authors. Therefore an extended dataset has been created using the QWS dataset which has additional important parameters which are not included in QWS dataset [8]. The technique presented in [9] has been employed to calculate value of the proposed parameters and their dependencies on one another. A parameter can depend on another parameter if there is a relationship between them. However, the issue of consideration is to observe the degree of these dependencies between parameters. These dependencies can be termed as low, high or medium.

## 3. SYSTEM ARCHITECTURE

The architectural diagram of the dataset creation system is given in Fig 1. In the figure, D shows the values of Performance or Availability as input. Firstly, the value of K is chosen as Standard Deviation (SD). After that the mathematical formulae is applied to calculate the values of required parameter P. Here, rel denotes the level of relationship (low or high). The process of deriving the parameters have been explained as stepwise procedure.

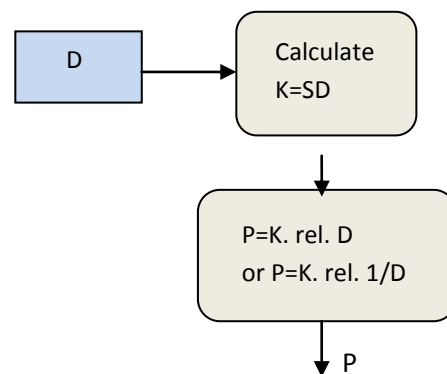


Fig 1

### 3.1 Steps of Dataset Creation

An extended dataset has been created using QWS dataset using the following steps. The dataset with 364 web services with four QoS (Quality of Services) parameters has been used to create the extended dataset. These services are classified into three classes on the basis of Highly Normalized Function (HNF). HNF is used to measure the quality ranking of a Web service based on quality metrics. The classification is based on the overall quality rating provided by HNF. Using HNF values obtained for each Web services, a classification scheme has been used for categorizing the web services.

The stepwise procedure has been given as follows:

#### Step 1:

The first step is to find the relationships(proportional or inversely proportional) amongst proposed additional QoS parameters and the given nine parameters in QWS Dataset. After deliberations, It was found that parameters security depends on performance and availability and hence these were chosen for calculating overall value of parameter security as without a certain level of performance and availability of a service, there is no meaning of secured service. Following were the observed relations between the three parameters:

$$\text{Security} \propto 1/\text{performance} \quad \text{OR}$$

$$\text{Security} = K1 \times 1/\text{performance}$$

$$\text{Scalability} \propto 1/\text{performance} \quad \text{OR}$$

$$\text{Scalability} = K1 \times 1/\text{performance}$$

$$\text{Interoperability} \propto \text{availability} \quad \text{OR}$$

$$\text{Interoperability} = K2 \times \text{Availability}$$

$$\text{Robustness} \propto 1/\text{performance} \quad \text{OR}$$

$$\text{Robustness} = K2 \times 1/\text{performance}$$

In all these equations, K1 and K2 are constants whose value would be calculated in subsequent steps taking appropriate assumptions as required.

#### Step 2:

Next step is to find what type of strength of relationship exists amongst them. And this relationship can be categorized in three categories which are Low, Medium and High.

$$[\text{Security}, \text{Performance}] \rightarrow \text{low}$$

$$[\text{Scalability}, \text{Performance}] \rightarrow \text{High}$$

$$[\text{Interoperability}, \text{Availability}] \rightarrow \text{Low}$$

$$[\text{Robustness}, \text{Performance}] \rightarrow \text{High}$$

These relations can be interpreted as security is related to performance with lesser degree than scalability with performance.

#### Step 3:

Some assumptions are to be taken to interpret domain values of low-medium-high in order to calculate value of the constants of K1 & K2 used in step1. As per the requirements, these values have been mapped to discrete values of low=0.3, medium=0.6 and High=0.9 in a scale of 0-1.

#### Step 4:

After making these assumptions, the value of constant K1 and K2 can be calculated. K1 is calculated from the values of performance parameter and K2 is calculated from the values of availability. The value of K is calculated by finding out the standard deviation of values of performance and availability. The value of K is chosen as Standard deviation (SD) because SD is a measure of the dispersion of a set of data from its mean and therefore it helps in reducing the outliers.

$$K = \sum_{i=1}^m \frac{X_i - u}{N - 1}$$

Where m=no of services, X=service parameter value, u=mean and N=total no of services.

#### Step 5:

After finding the value of constants K1 & K2, utilizing the relations between parameters in step1, the values of the extended parameters can be calculated as per the strength of the relationship as:

$$\text{Security} = K1 \times 30/\text{performance}$$

$$\text{Scalability} = K1 \times 90/\text{performance}$$

$$\text{Interoperability} = K2 \times 30 \times \text{Availability}$$

$$\text{Robustness} = K2 \times 90/\text{performance}$$

At the end of this step, a set of 364 web services with corresponding additional four parameters along with calculated values evolves.

#### Step 6:

The next step is to normalize the parameter values by applying a normalization Function. The HNF (Highly Normalized Function) has been used which is used for uniform distribution of values.

Here A represents the matrix representation of QoS values where  $q_{ij}$  represents the  $j^{\text{th}}$  QoS value for  $i^{\text{th}}$  web service

$$A = \begin{bmatrix} q_{1,1} & \cdots & q_{1,j} \\ \vdots & \ddots & \vdots \\ q_{i,1} & \cdots & q_{i,j} \end{bmatrix}$$

This represents the normalized value of  $i^{\text{th}}$  web service based on j QoS parameters

$$\text{Nor}_{(j)} = \sum_{m=1}^i q_{m,j}$$

Normalizing the Nor value for  $i^{\text{th}}$  web service on j parameter

$$\text{NorDiv}_{i,j} = \frac{q_{i,j}}{\max(\text{Nor}_{(j)})}$$

The following matrix represents the final normalized values of QoS parameters for all the web services.

$$A' = \begin{bmatrix} \text{NorDiv}_{1,1} & \cdots & \text{NorDiv}_{1,j} \\ \vdots & \ddots & \vdots \\ \text{NorDiv}_{i,1} & \cdots & \text{NorDiv}_{i,j} \end{bmatrix}$$

After Normalization of the parameters, the mean of all the parameter is taken. Then it is subtracted from all the values of individual parameters to reduce the effect of mean. It will reduce the effect of difference between low value and high

value of any parameter. All parameter values are therefore considered equally. After this the calculated HNF has been used to categorize the services where:

$$\text{HNF}(ws_i) = \sum_{i=1}^m (\text{NorDiv}_{i,j} - \text{mean})$$

Where  $m$  represents the number of web services

#### Step 7:

Further, the set of services has been classified into three classes or three categories. The classes range is given in the table 1 below.

**Table 1. Classes of Services**

Class	HNF Value
1	$\geq 87$
2	86 to 70
3	$\leq 69$

An NfPD dataset has been prepared with total 6 parameters. A sample example is given in table 2.

**Table 2. Example of prepared dataset**

Service	Security	Scalability	Inter-operability	Robustness	HNF	Class
ErrorMailer	6.38	19.16	8.31	19.16	99	1
XigniteNews	35.08	59.31	50.44	90.18	73	2
getJoke	15.42	46.25	31.07	46.25	51	3

## 4. RESULTS

Using the steps given in section 3, a Non-functional parameter Dataset has been prepared which consists of 364 real service having actual values with 6 parameters. These are the important parameters left in earlier proposals. In this proposal a methodology has been designed to calculate the additional parameters. Four additional parameters: Security, scalability, robustness and interoperability have been derived using statistical techniques to harness the various degrees of relationships between various QoS Parameters. The designed NfPD dataset is altogether numeric. In the nutshell, this paper has not only extended the existing QWS dataset but also proposed a classification technique based on overall QoS values of web services using HNF. This proposal therefore can help in improving the QoS based service offering to a client.

**Table 3 Dataset Details**

Samples	Attributes	Type	Missing	Classes
364	6	Numeric	0	3

## 5. CONCLUSIONS

In this paper, a method has been proposed to generate a synthetic web service dataset which is named as Non-functional Parameters Dataset (NfPD). This dataset consists of four important QoS parameters which were left in QWS Dataset. This dataset can be used in data mining research and development. Our dataset can be used in the data mining where numeric dataset is required which is the only limitation of NfPD dataset.

## 6. REFERENCES

- [1]Jaideep Roy and Anupama Ramanujan, "Understanding Web Services", 2001 IEEE
- [2]XML <http://www.w3schools.com/xml/default.asp>
- [3]Web Services [www.w3schools.com/webservices/ws\\_intro.asp](http://www.w3schools.com/webservices/ws_intro.asp)
- [4] SOAP Version 1.2 Part 0: Primer W3C Recommendation 24 June 2003, Available at: <http://www.w3.org/TR/soap12-part0/>, Accessed on 06/04/2013.
- [5] Universal Description Discovery and Integration (UDDI) Version 3.0.2 at: [http://www.uddi.org/pubs/uddi\\_v3.htm](http://www.uddi.org/pubs/uddi_v3.htm),
- [6] Data Mining, [en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining)
- [7] Matthias Seeger, "Learning with labeled and unlabeled data", Institute for Adaptive and Neural Computation, University of Edinburgh
- [8] Eyhab AL-Masri and Qusay H. Mahmoud, "Towards Quality-Driven Web Service Discovery", University of Guelph, 2008 IEEE
- [9]Service Oriented Architecture [http://java.sun.com/developer/Books/j2ee/jwsa/JWSA\\_C\\_H02.pdf](http://java.sun.com/developer/Books/j2ee/jwsa/JWSA_C_H02.pdf) [www.service-architecture.com/web-services/articles/](http://www.service-architecture.com/web-services/articles/)
- [10] Jeffrey W. Seifert, "Data Mining: An Overview", Analyst in Information Science and Technology Policy Resources, Science, and Industry Division
- [11] Vandan Tewari, Nirmal Dagdee and Aruna Tiwari. Article: User Oriented Web Services Discovery based on QoS Parameters in Multiple Registries. *International Journal of Computer Applications* 46(24):8-12, May 2012. Published by Foundation of Computer Science, New York, USA
- [12] Al-Masri, E., and Mahmoud, Q. H., "Discovering the best web service", (poster) 16th International Conference on World Wide Web (WWW), 2007, pp. 1257-1258
- [13] Al-Masri, E., and Mahmoud, Q. H., "QoS-based Discovery and Ranking of Web Services", IEEE 16th International Conference on Computer Communications and Networks (ICCCN), 2007, pp. 529-534.
- [14] Hassan Reza, Dan Jurgens, Jamie White, Jason Anderson, and Jay Peterson, "An Architectural Design Selection Tool Based on Design Tactics, Scenarios and Nonfunctional Requirements", University of North Dakota, EPSCoR NASA grant #NCC5-582