International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 2 No. 3 Mar. 2013, pp. 230-237

http://www.issr-journals.org/ijias/

# Measuring Semantic Similarity in Grids Using Ontology

***karamollah Bagherifard[1-2], Mehrbakhsh Nilashi[1], Othman Ibrahim[1], Norafida Ithnin[1], and Lasisi Ayodele Nojeem[3]***

[1]Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia,
Skudai, Johor, Malaysia

[2]Department of Computer Engineering,
Islamic Azad University, Yasooj branch,
Yasooj, Iran

[3]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia



**ABSTRACT:** Grid computing, a new and broad area of research, aims at sharing available information and resources through the use of computers over the network. To use the new applications of grid, it is necessary to adapt the modern software components and assembled information resources in a flexible format. Web services incorporate the necessary capabilities in achieving this goal called grid services. Due to the exponentially increasing amount of data, documents, resources and services available on the web, finding an acceptable agreement between the user and the abilities of web or grid service as well as forming an appropriate composition of service components for performing requested operation are critical issues. Measuring the similarity of services is an important and valuable solution that is used in some practical reasoning such as replacement of a service with another and combination of services and applications. Also, because the measuring the service similarity needs an appropriate semantic model, therefore, in this paper a semantic model based on OWL ontology language for services is presented and thus, similarity measure is provided. We find a semantic model for services and then provide a method for measuring the similarity between two services. A mathematical model for solving given problems is also proposed. The results evaluated by F1 measure obviously show the improvement of accuracy against previous method.

**KEYWORDS:** Grid, Web Services, Similarity, Semantic Web.

## 1 INTRODUCTION

The grid can be considered as a network layer services that allow users to access the set of distributed and computing resources, applications, and data resource. A grid service causes the entire network to be viewed as a seamless information processing system accessed by users at any situation. The concept of grid services emanated in 1990 providing solutions for resource sharing with high performance that deals with huge amount of computing of large data [2]. To use the new applications of grid, it is necessary to adapt the modern software components and assembled information resources in a flexible format. With respect to this and changes in procedure of a set of protocols, grid has been transformed to an application and a service-oriented method. Open Grid Service Architecture (OGSA) incorporates the grid techniques and web services. A grid service, in brief, is a Web service that follows a specific set of rules (institutions and intermediaries) that define how the user interaction with the Grid services [14].

With exponentially increasing amount of data, documents, resources and services available on the web, finding an acceptable agreement between the user and the abilities of web or grid service as well as forming an appropriate composition of service components for performing requested operation are critical issues because humans are not able to provide effective and efficient means for the description of services, components and objects that are available on the web [16].

Measuring the similarity of services is an important issue in many applications such as service discovery, service composition and recommendation. Due to the increasing number of services, measuring the service similarity needs an appropriate semantic model as proposed in this paper. We proposed a semantic model based on OWL ontology and in regard to this model, we use semantic similarity for calculating services similarity [17, 1]. The remainder of the paper is structured as follows:

Semantic model in services is described in the Section 2 for further discussion. Also, semantic similarity methods are presented in Section 3 and afterward, using semantic services models and semantic similarity methods in Section 4. A solution to the described problems is provided in Section 5 and finally, evaluation of proposed method and conclusion is given in Section 6 and 7 respectively.

## 2 SEMANTIC SERVICE

The question that arises here is why the grid needs semantic? The answer is, first, the development of the grid without the use of Semantic Web technologies is thus reducing transparency for users [1], because interpreting and managing the huge volume of resources by human (users) is not easy. Using semantic web technology in interpreting the resources, users' effort and attempts will be reduced and the use of resources will be effective and efficient [12]. The languages that describe the service such as WSDL; only consider the Syntactic description of the service [8] and the information about what the services performs is not provided to the user. Therefore, the user has to provide additional explanations about the service.We have presented a semantic model for service and applied OWL ontology languages to model services. In this system, each Web service is a class and relations between them are modeled based on the OWL ontology tags.

Classes of the tags ontology are displayed based on OWL [18]. Each class may include a subclass by OWL and Subclass tag is displayed in OWL language. In these models, the class that includes sub-class, represents a combination of services. Each class has a feature name that specifies the service name. In this paper, six properties are considered for each service that through features, definitions in the OWL ontology language is modeled [5]. A service has its own specific non-functional descriptions such as location, characteristics and so on. All these features are included into the characteristic called metadata. In addition, we consider a special feature called the usage which represents applications of service. For each service, we consider a feature called "IS-A" and last feature is reference that represent the resources in which will be consumed by web services or grid. Figure 1 shows the features of a service in the ontology.

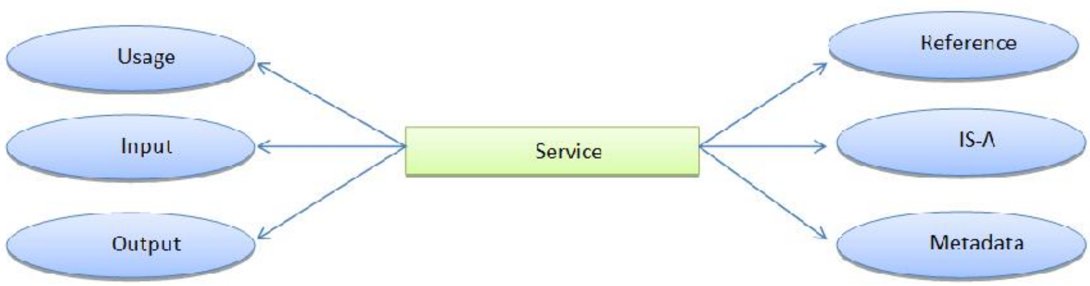


*Fig. 1. Features of Service*

Input features indicate the type of input data that are necessary condition for running the previous service. Output features indicate the type of output data and results service and "IS-A" feature reflects the service that is the current service components. This feature is important in several respects [6]. Through that, sub-services of compound service can be specified. Its main use is to determine the relation similarity among services between services that in next section further studies will be investigated.

## 3 Concept of Similarity Measuring

After defining a semantic model for services, it is necessary to present a method for measuring the similarity between two services. Therefore, in this section some of similarity measure is introduced, two methods are considered for semantic similarity. At the end of this section, a vector similarity measure is also expressed.

### 3.1 Similarity Measuring Based on the Classification

One of the existing relationship between the concepts is "IS-A" relationship. Using this relationship, we are able to classify the related concepts. If C be a set of concepts, therefore, the classification of concepts is defined as $\tau(C,\leq)$ that $C \leq C'$ means $C\ IS-A\ C'$ [19, 7].

In this method, two concepts as one ontology should be placed where it tends to be grouped in one class. If two concepts are different in the two ontologies, they are therefore merged and form a unified ontology [3]. The similarity between two concepts is defined as [7]:

$$\sigma(C,C') = 1 - \delta(C,C') \tag{1}$$

Where $\delta(C,C')$ indicates weighted distance between two concepts. In classification for any concepts C, the weight W(c) exists and $cpp(c_1,c_2)$ is the weight of common father for two concepts $c_1$ and $c_2$.With considering these definitions, the value of $\delta(c,c')$ is calculated as :

$$\delta(c_1,c_2) = \left[w(ccp(c_1,c_1) - w(c_1)\right] + w\left[ccp(c_1,c_2) - w(c_2)\right] \tag{2}$$

The concepts weight of classification is calculated by [19,7] :

$$w(n) = \frac{1}{k^{l(N)+1}} \tag{3}$$

In this equation, $L(N)$ indicates the length of path to the node N in the classification tree and K denotes a pre-defined value greater than 1 that here value 2 is considered .This model has two main properties: (1). in high levels, the concept differences are more than existing differences in the lower levels. (2).the distance between two concepts is more than difference between child and parent. For example consider the Figure 2, in this Figure we want to calculate the similarity degree between two concepts car and truck that following steps are performed as shown in Figure 3.

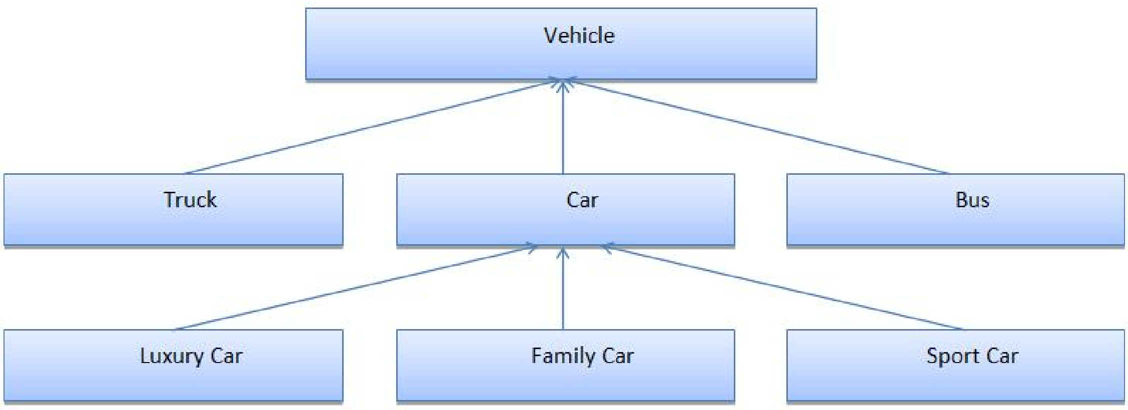


*Fig. 2. An example of concept classification*

$$\omega(Vehicle) = 0.5$$
$$\omega(Truck) = 0.25$$
$$\omega(SportCar) = 0.125$$
$$\delta(Truck, Vehicle) = 0.5 - 0.25 = 0.25$$
$$\delta(SportCar, Vehicle) = 0.5 - 0.125 = 0.375$$
$$\delta(Truck, SportCar) = 0.25 + 0.375 = 0.625$$
$$\delta(Truck, SportCar) = 1 - 0.625 = 0.375$$

***Fig. 3. Steps for calculating similarity degree between two concepts car and truck***

### 3.2 MEASURING THE SIMILARITY BASED ON FEATURES

In the previous methods for calculating the semantic similarity of concepts, only the classification structure of concepts was considered and other concepts and features were overlooked. Each concept has a number of features and a set of object features [6].

Different characteristics of the concept are identified by features. For example, consider two of the Person of the Father. One concept of Person [6,7] can have a property of a data type called the Father and Father concept can have a property as SubClassOf of a relation related to this concept. Father also has a relationship as HasChild with Person that is defined as following:

F (Person) = {(type, class), (HasName, String)}

F(Faher )={ (type, class ), (HasHame,String ) ,

(SubClassOf , Peron ) , (HasChild , Person ) }

In the example above, because the Father is one SubClass of Person then inherit the HasChild property .Two more general properties are important rather than a particular property of a concept. Similarity between two concepts $c_1$ and $c_2$ can be defined as follows [6].

$$\sigma(c_1, c_2) = \frac{2 \times |f(c_1) \cap f(c_2)|}{|f(c_1) \cap f(c_2)| + |f(c_1) \cup f(c_2)|} \qquad (3)$$

The success of these measures depends on the degree of context properties. In most of the current ontologies only relationship between concepts in ontology is defined and rest relationships are ignored. In this type of ontology similarity, measuring between two concepts based on feature cannot be useful, and the result is often not acceptable.

**Vector model.** Words or concepts indicators in this model (e.g., questions and documents), are defined as weighted. The matching degree of two vectors denotes the similarity degree between them. Two concepts are considered in the t dimension space and similarity between text and question is defined by similarity between $p_i$ and $q_j$ vectors.

**Dice measuring.** This method is used for measuring similarity between two concepts defined as vectors. The equation 5 for this measuring method is formulated as

$$sim_{dice}(d_s, q_i) = \frac{2 \times [\sum_{k=1}^{n} (w_{s,k} \times w_{i,k})]}{\sum_{k=1}^{n} w_{s,k} + \sum_{k=1}^{n} w_{i,k}} \qquad (4)$$

## 4 SERVICE SIMILARITY

Similarity between services can be looked at from several aspects that each of these similarities in respect to the functional problem is useful. In the following we will address three aspects of the service similarity [9, 10].

$$\Pi(c_1,c_2)=\begin{cases}\dfrac{\max_{c1\in C1,c2\in C2}\{\Pi(C1-c1,C2-c2)-\sigma(c1,c2)\}}{\min\{|C1|,|C2|\}} & C1\neq\Phi, C2\neq\Phi\\ 0 & C1=\Phi\vee C2=\Phi\end{cases} \quad (5)$$

### 4.1 SIMILARITY BETWEEN INPUTS OF TWO SERVICES AND THEIR OUTPUTS

Similarities between the inputs and outputs of the two services are an appropriate method for measuring the similarity of web services. The inputs and outputs of a service is in fact a collection of elements, we consider each element as a concept. In principle, for measuring of similarity between inputs in two services, similarity between two set of concepts must be calculated [9]. Ganjisaffar [1] developed a model for measuring the similarity between a set of concepts; in this model semantic similarity measure methods (classification and feature similarity measuring) are used (Equation 5).

Similarity between two services is defined as sum of inputs and outputs similarity in two services as shown in equation 6.

$$F1(S_1,S_2)=\frac{\Pi(I1,I2)+(O1,O2)}{2} \quad (6)$$

### 4.2 METADATA SIMILARITY BETWEEN TWO SERVICES

Based on the above methods, similarity between two services is calculated through their input and output and content services have been unnoticed. In some application cases, the contents and concept of the services are also considered. For this reason, we present the similarity measuring among metadata services. A reasonable method to calculate similarities between metadata is the vector model that we introduced in this paper. Metadata for each service is considered as two vectors $W_1$ and $W_2$ and similarity of two services is identified from similarity of metadata [20, 21].

$$F2(S_1,S_2)=sim_{Dice}(W_1,W_2) \quad (7)$$

### 4.3 FUNCTIONAL SIMILARITY OF TWO SERVICES

As mentioned earlier, for each service, a feature called usage is considered and through this feature it is possible to identify the functionality of any services. In this method, the similarity of services is measured based on their application. Applications can be considered as a concept and the similarities between them are calculated through classification similarity measure [4, 9, 11].

$$F3(S_1,S_2)=\sigma(S_1,S_2) \quad (8)$$

## 5 FUNCTIONAL CASES: COMBINING SERVICES

Once a request by available atomic services is not fulfilled, it is possible to fulfill this request using the proper integration and composition of existing services. The process of gathering atomic services to create an integrated and coordinated combination set is called combining services that fulfills the larger and more complex purpose than what is done by the individual atomic services [13, 15]. Indeed, the possibility of integrating and combining the services by different organizations for fulfilling the user request is one of the factors that services become attractive.

### 5.1 SELECTING APPROPRIATE SERVICE FOR COMBINING

When two services are combined, the input of second service is equivalent to the output of first service. Therefore, inputs and outputs data types and application of them must be compatible to each other [2,13]. Here the similarity between inputs and outputs is not necessary but the similarities between the output of the first service and an input of second service is important. Therefore, equation F1 is defined as follows with $S_1$ , $S_2$ ,$O_1$ and $I_2$ denoting first service, second service, output of first service and input of second service, respectively

$$F1(S_1, S_2) = \Pi(O_1, I_2) \qquad (9)$$

In addition to the inputs and outputs similarities, the conceptual and practical similarity of service is also considered. Using the equation 10, the two services are combined if the high similarity between them is obtained.

$$similarity(S_1, S_2) = W1 * F1(S_1, S_2) + W2 * F2(S_1, S_2) + W3 * F3(S_1, S_2) \qquad (10)$$
$$W1 + W2 + W3 = 1$$

### 5.2 Selecting appropriate service for replacement

When running the service, the sub service failed and therefore, this causes the halt of running in whole service. Our solution is finding and replacing the similar service with failed service. For measuring similarity of services in this functional case, three combinations of functions F1, F2 and F3 are offered.

$$similarity(S_1, S_2) = W1 * F1(S_1, S_2) + W2 * F2(S_1, S_2) + W3 * F3(S_1, S_2) \qquad (11)$$
$$W1 + W2 + W3 = 1$$

## 6 Evaluation

For evaluating and demonstrating accuracy of the proposed method, we used the F1 and the evaluating results are given in the following(see Figure 4 and 5).

$$Precision = \frac{\text{size of hit set}}{\text{size of top-N set}}$$

$$Recall = \frac{\text{size of hit set}}{\text{size of test set}} \qquad (12)$$

$$F1 = \frac{2 * Recall * Percision}{(Recall + Percision)}$$

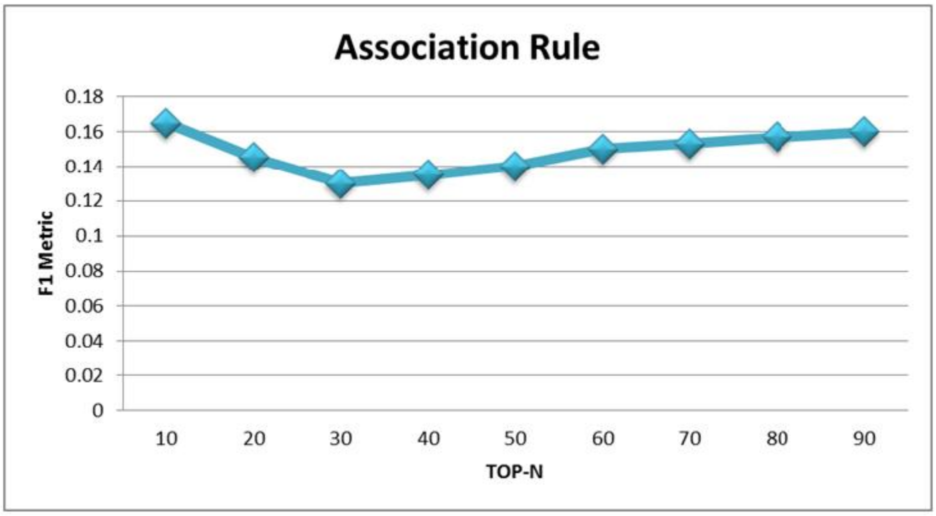


*Fig. 4. F1 result for similarity measuring in previous method*

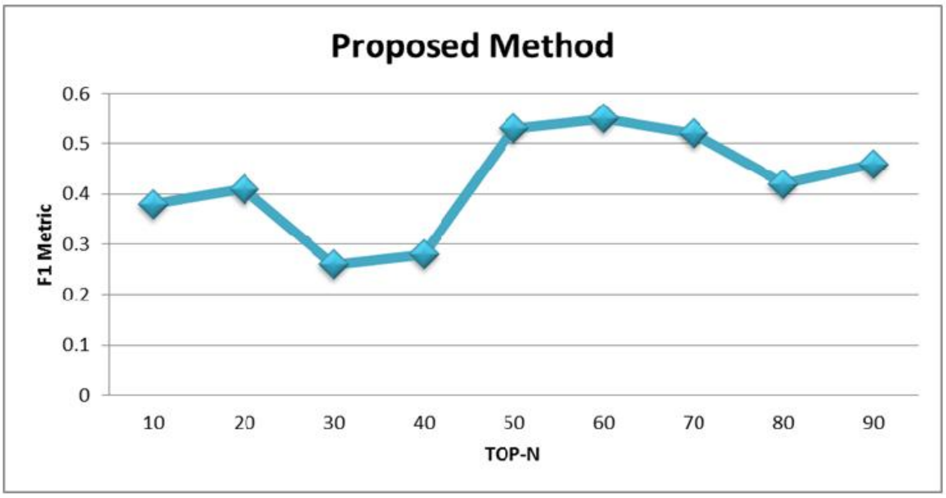


***Fig. 5. F1 result for similarity measuring in proposed method***

## 7 CONCLUSION

In this paper, methods for similarity measure among services were presented. The main advantage was incorporating concept in the proposed methods and similarity of services is calculated based on concept. Through measuring the similarity of services, models for combining and replacing the services also were presented and especially these models consider the issue of application of services. This consideration causes the accuracy enhancement in selecting appropriate service. The results evaluated by F1 measure obviously show the improvement of accuracy against previous method. Similarity measuring was calculated between 0.13 and 0.17 for the previous method while in the proposed method it was calculated between 0.26 and 0.55, thus showing an improvement of the accuracy.

## REFERENCES


[1] Ganjisaffar, Y., Abollhasani , H., Neshati, M., and Jamali , M., “A Similarity Measure for OWL-S Annotated Web Services”, Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings), 2006, Page(s): 621 - 624.

[2] Xiaofeng Du, William Song, Malcolm Munro, “Service Composition in the Context of Grid”, 2006.

[3] G. Stumme and A. Maedche, “Ontology merging for federated ontologies on the semantic web,” in Proceedings of the International Workshop for Foundations of Models for Information Integration (FMII-2001), Sept. 2001.

[4] S. Melnik, H. Garcia-Molina, and E. Rahm, “Similarity flooding: A versatile graph matching algorithm and its applications to schema matching,” in *Proceedings of the 18th International Conference on Data Engineering*. San Jose, CA: IEEE Computer Society, Feb. 2002.

[5] Martin, D., Paolucci, M., Mcllraith, S., Burstein, M., McDermott, D., McGuinness, D., Parsia, B., Payne, T., Sabou, M., Solanki, M., Srinivasan, N., and Sycara, K., “Bringing Semantics to Web Services: The OWL-S Approach”, presented at First International Workshop on Semantic Web Services and Web Process Composition (SWSWPC), San Diego, California, USA, 2004.

[6] A. Tversky, “Features of similarity,” *Psycological Review*, vol. 84, no. 4, pp. 327–352, 1977.

[7] P. Raftopoulou and E. Petrakis, “Semantic similarity measures: A comparison study,” Technical University of Crete, Department of Electronic and Computer Engineering, Tech. Rep. TR-TUC-ISL-04-2005, Jan. 2005.

[8] Foster, I. and Kesselman, C., The Grid: Blueprint for a New Computing Infrastructure, Publisher: Morgan Kaufmann Inc, 1999. ISBN: 1-555860-475-8.

[9] Sung Lee, Taowei David Wang, Nada Hashmi, Michael P. Cummings, “Bio-STEER: A Semantic Web workflow tool for Grid computing in the life sciences,” Future Generation Computer Systems, Volume 23, Issue 3, March 2007, Pages 497-509.

[10] Kenneth J. Turner, Koon Leai Larry Tan, “Rigorous development of composite grid services”, Journal of Network and Computer Applications, Volume 35, Issue 4, July 2012, Pages 1304-1316.

[11] Tai-hoon Kim, Wai-chi Fang, “Special section: Grid/distributed computing systems security”, Future Generation Computer Systems, Volume 25, Issue 3, March 2009, Page 351.

[12] Simone A. Ludwig, S.M.S. Reyhani, “Semantic approach to service discovery in a Grid environment”, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 4, Issue 1, January 2006, Pages 1-13.
[13] H. Kevser Sunercan, M. Nedim Alpdemir, Nihan Kesim Cicekli, “A systematic approach to the integration of overlapping partitions in service-oriented data grids”, Future Generation Computer Systems, Volume 27, Issue 6, June 2011, Pages 667-680.
[14] Antonios Litke, Kleopatra Konstanteli, Vassiliki Andronikou, Sotirios Chatzis, Theodora Varvarigou, “Managing service level agreement contracts in OGSA-based Grids”, Future Generation Computer Systems, Volume 24, Issue 4, April 2008, Pages 245-258.
[15] Fei You, Qingxi Hu, Yuan Yao, Gaochun Xu, Minglun Fang, “study on web service matching and composition based on ontology,” Computer Science and Information Engineering, 2009 WRI World Congress Volume: 4.
[16] Jiehan Zhou, Pakkala, D., Perala, J., Niemela, E., Riekki, J., Ylianttila, M., “Dependency-aware Service Oriented Architecture and Service Composition”, Web Services, 2007. ICWS 2007. IEEE International Conference.
[17] Min Liu, Weiming Shen, Qi Hao, Junwei Yan, “An weighted ontology-based semantic similarity algorithm for web service,” Expert Systems with Applications, Volume 36, Issue 10, December 2009, Pages 12480-12490.
[18] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. The Description Logic Handbook. Cambridge University Press, 2002.
[19] David Sánchez, Montserrat Batet, David Isern, Aida Valls, “Ontology-based semantic similarity: A new feature-based approach”, Expert Systems with Applications, Volume 39, Issue 9, July 2012, Pages 7718-7728.
[20] Xin Dong, Alon Halevy, Jayant Madhavan, Ema Nemes, Jun Zhang, “Similarity Search for Web Services”, *Proceedings 2004 VLDB Conference*, *2004*, *Pages 372-383.*
[21] David Loshin, “10 - Metadata and Data Standards”, The Practitioner's Guide to Data Quality Improvement, 2011, Pages 167-189.