A Better Approach for Conceptual Readability of WSDL

Kashif Sohail Abid*, Asif Sohail Abid† and M. Mohsen Ansari‡
*School of Electrical Engineering and Computer Sciences,
National University of Sciences and Technology,
Islamabad, Pakistan.
Email: 10msitksohail@seecs.edu.pk
†College of Electrical and Mechanical Engineering,
National University of Sciences and Technology,
Islamabad, Pakistan.
Email:asifsohail83@gmail.com
‡Faculty of Computing,
Riphah International Unierysity
Islamabad, Pakistan.
Email:mohsen.ansari@riphah.edu.pk

***Abstract*—Issues that concerns with the inter-operability on a heterogeneous environment can easily be address using the flexible platform of Service Oriented Architecture (SOA). Web service is an implementation and modeling of Service Oriented Architecture (SOA). Web service description language (WSDL) is a standard describing a web service in XML form. This Description can be categorized in two parts i.e. structural and non-structural. The readability of a web service helps the consumer to understand it easily, it is suggested to provide sufficient details about functionality scope and limitation of scope in WSDL, so that it can easily be understandable. Readability depends upon interaction of two variables i.e. text and reader. The maximum details about a web service could lead to it's reproduction by business competitor, and it may helps in maximizing vulnerabilities in it. This paper focuses on a technique for computing readability index by a detail analysis of WSDL document. This readability index obtain using this approach helps the producer of a web service to adjust readability, so that it can easily be understandable by consumer. The better readability index can also leads the provider to a better service discovery. To calculate Readability Index, extraction of WSDL file components was performed. After extraction of key concepts, they were mapped with the Domain Ontology. The words that were not mapped in the ontology, synonyms are employed by consulting the WordNet. Final readability was obtained using Simplified Dale Chall readability index (DaCw). The Web Service Readability can be measure more precisely by considering words that were not found in the mapping process.**

***Keywords*-Web Services, Readability, WSDL**

I. Introduction

Web Service is one of the promising technologies that enable implementation and modeling of SOA, it has attained the focus of both academia and industry peoples resulting in an increase in number over internet [1]. WSDL (Web Service Description Language) is a standard based on XML that describes the web services using different components. The description of a Web Service can be divided into two parts i.e. structural and non-structural. Structural part <types>, <message>, and <portType>holds the information for exchanging the data, capabilities and port which a web service is using [2][3]. It helps the consumer to understand what a web service looks like? The Documentation element of WSDL document holds the description of the web service in the natural language form provided if any[4]. The information contains in WSDL are given by the provider of a Web Service. The description of a web service plays a vital role in understanding the functionality and usability of the web service. The readability of a web service is a concern to a provider, as it helps the consumer of the web service to understand it with an ease [5]. It is advised to use the meaningful names of operations, interfaces and messages. The documentation element of a large number of WSDL document are often found missing or contains very few words, which results in difficulty of understanding the web service by the consumer [6]. It is also suggested to provide the sufficient details about the functionality, scope and limitations in WSDL.The readability is dependent on two variables text and reader, interaction of these two variables helps in determining the readability of a material for any reader [7].

On the other side, providing maximum details about a web service can leads to understanding and reproduction of service by business competitors. Security attackers can use this information and guessing remaining required as a help in launching the attacks like information espionage, command injection, and denial of service. So, sharing of such information at a maximum level can also lead in severe attacks. The service discovery depends on readability of a web service, maximum readability can adds in vulnerability of web service. Considering it as a trade-off, the provider of web service should design the description of a service in such a way that minimize the attack vector and maximize the readability of it.

The researchers have also focused on the document readability with respects to the concepts used in it, by retrieving

978-1-4799-8688-0/15 $31.00 © 2015 IEEE
DOI 10.1109/BigMM.2015.52

the information from a specific domain. The readability of web service can be determine using different readability measurement approaches, and it can be utilize by the provider for developing it in a manner that can be helpful for the consumer of a web service.

This paper focuses on a technique for finding the readability index by a details analysis of WSDL document based on different layers of components. It reads the WSDL documents, extracts the specified information, and converts them to the tokens after removing the stop words from it. After this phase it utilizes these token to determine different key concepts presence in the Document by using Domain Ontology provided. The ontology is the representation of concepts of a domain, it's attributes and relationship that uses Web Ontology Language (OWL) an XML based language to represent them. To provide a thorough analysis, WordNet [9] is also attached with the matcher. Conceptual Readability index calculated in this approach is based on the conceptual readability model provided by Yan et al[10].

The rest of paper is organized as, Section II contains the related work, Section III contains the discussion on Conceptual Readability Model, Section IV discusses the Architecture of Proposed Scheme, Section V contains the Implementation and Results, and Section VI presents the Conclusion.

II. Related Work

The readability of a document is a value that is calculated using some mathematical formula. It is a measure of difficulty level for a text. A text may be harder for a particular reader, whereas the same text may be simple to understand for other reader. The readability depends upon two variables of reader and text. The interaction of these two variables determines the level of difficulty. Readability is not precise measure but helps in understanding complexity of a text for a reader. The interactive view of readability is also focused by different other researchers. The assessment of this readability has its purpose to effect a best match between two said variables, interaction and purpose of reading is vital to find the optimal difficulty level. The conceptual readability of a text depends upon different factors like density of concepts, coherence, and organization of the text.

Pananya et al. [8] works on proposing the approach for the measurement if the readability index of a WSDL document. The provider can reduce or increase the web service readability index to achieve the desired readability, so as to minimize the attacks issue on the web service. Their approach provides a readability measurement based on the keywords or terms in service domain knowledge. Its reads the WSDL document, parse the contents, and check these keywords for the presence in the domain concepts. It involves different mathematical steps for calculations of density of concepts, Document Cohesion (DC), and Document Scope (DS). To calculate the readability index it uses the Simplified Dale Chall readability index (DaCw) [11].

As already mentioned that the non-structured part of the Web Service contain the description in Natural Language. The provider may have used the words that are present in the Domain ontology but in a synonym form. On the contrary there may be a scenario were the Domain ontology refuse to provide a match for a term from the WSDL document. The approach presented in this paper uses a scheme that uses WordNet [9] at different steps to address the solutions of the highlighted issues.

III. Conceptual Readability Model

A. Document Scope (DS)

Document Scope is defined as the coverage of domain concepts in a document. The presence of maximum domain concepts in a documents leads to it's less readability for an average user. A document will have more difficulty if domain terms present in it are deeper in concept hierarchy. The scope of WSDL document (Scope (di)) can be computed by equation 1.

$$scope(d_i) = e^{-(\sum_{i=1}^{n} depth(c_i))} \tag{1}$$

Depth (ci) is the depth of a particular domain concept from the WSDL document in the concept hierarchy. The time complexity of the scope based readability measurement is O(mn), where m is the number of WSDL documents.

B. Document Cohesion (DC)

The document cohesion is another feature of the document and is also referred as semantic relatedness. The degree of cohesiveness in terms defines the relatedness property of the document. When the document terms are map to domain ontology (concepts) then distance between the matched terms defines cohesion between them, this can be computed by equation 2.

$$Cohesion(d_i) = \frac{\sum_{i,j=1}^{n} Sim(c_i, c_j)}{Numberof Association} \tag{2}$$

Where n is the number of domain concepts in the WSDL document di, and n >1, i<j. the number of association is the total number of association between the domain concepts and can computed by using equation (3)

$$Numberof Association = \frac{n(n-1)}{2} \tag{3}$$

The Sim(c_i, c_j) is the similarity of the two concepts c_i and c_j. It is also referred as semantic similarity, and it can be calculated by the equation 4.

$$Sim(c_i, c_j) = -log\frac{len(c_i, c_j)}{2D} \tag{4}$$

Where D is the maximum tree depth, and $len(c_i, c_j)$ is the distance between two concepts in the Domain hierarchy.

C. Simplified Dale-Challś Readability Index (DaCw)

The Dale-Challs readability index score (DaC) is mathematically presented as

$$DaC(d_i) = (0.0496 * AvgSL) + (0.1579 * PDW) + 3.6365 \tag{5}$$

Where AvgSL is the average length of the sentence in the document di , and PDW is the percentage of difficult word calculated from the list of Dale-Chall. As the focus of this research is on the word level readability, so we are left with the equation 6.

$$DaCw(d_i) = PDW \quad (6)$$

The Dale-Challs provides the list of 3,000 familiar words and considers those words for making the hypothesis for difficulty of the document.

D. Concept-Based Readability

Concept-Based Readability (CRS) defines the overall readability of the WSDL document, greater the CRS the more difficult a document will be to read. CRS can be calculated mathematically as,

$$CRS(d_i) = Scope(d_i) + Cohesion(d_i) + DaCw(d_i)^{-1} \quad (7)$$

The scope of the document is added with the cohesion and inverse of simplified Dale-Challs is added to achieve Concept-Based Readability of the WSDL document.

IV. Architecture of Proposed Scheme

The proposed scheme works in different steps to achieve the Concept-based readability results. These steps are shown in the figure-1. The user interface is input screen where user tells the path of the WSDL document for the Concept-Based Readability, this step also accept the parameter of whole or partial document. In this step the extraction of the terms is performed from the WSDL document. This extraction is made either partially or as a whole. Removal of stop words from the extracted text of WSDL and maintaining the terms in a data structure is performed in this step. These terms are used for the next step of the proposed scheme. The matcher component performs two main functions.Firstly it matches the presence of the concept extracted from the wsdl document into the domain ontology provided. Secondly, the proposed scheme also takes into consideration the case where no match is found, to address this issue the WordNet [9] is attached in this layer. If a term is found unmatched in the domain ontology then the synonym of the term is taken on its behalf. As discussed earlier that to calculate the Concept-Based Readability the Document Scope, Document Cohesion, and Simplified Dale-Challs index is required. These calculations are performed in this step, the list of 3,000 common words matched to the terms that were extracted from the WSDL document. The overall readability is calculated from the equation 7, and result is shown to the service provider.

V. Implementation and Results

To perform the readability assessment, prototype system is implemented in Java NetBeans IDE. The WSDL document of Amazon, eBay, KonaKart Enterprise , and PayPal were selected for the readability assessment.Observing the space limitation, an example of Amazon's WSDL is presented here.

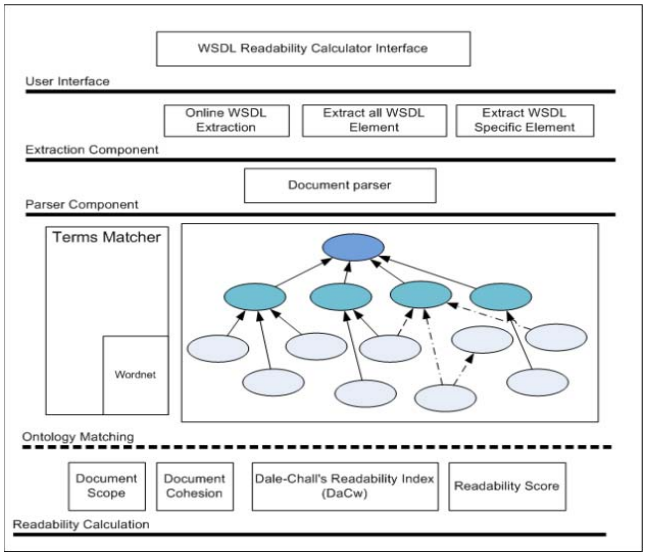

Fig. 1. Architecture Diagram

It will help in explaining the functionality of the proposed system.

In first step WSDL file is given to the proposed system, it can select the partial or complete document as specified. The proposed system extraction WSDL elements from the WSDL document. In this step removal of stop words is also performed from the unstructured document. The extracted terms from the WSDL were 1007, where some of the domain terms extracted from the WSDL are shown in table 1.

TABLE I
Some Terms Extracted from Amazon's WSDL

price	delivery	charge	week	promise	quantity
payment	mode	offering	opening	specification	node

At this stage the duplicate words are removed, and conversion to singular form is performed. So in this stage preparing input for the mapping stage is being done.

In this step the final remaining words are mapped into GoodRelation concept hierarchy. This mapping helps to obtain the parameters that will be helpful in calculation of conceptual readability. If a word results in "not found" during the mapping process then its synonym is extracted from the attached WordNet. This step is important when the unstructured WSDL part is under consideration during calculation conceptual readability. For example the provider has used a word "truck" in the documentation and it is not present in the concept hierarchy whereas its synonym "vehicle" exists in the concepts. so there is a need to consider this word during readability calculation, also it is required for enriching the domain ontology. 139 words were matched in the domain concepts, table 2 shows some of them.

To calculate document scope, depth of these concepts is obtain. The depth of some of the concepts is shown in the

TABLE II
SOME TERMS THAT WERE FOUND IN THE DOMAIN ONTOLOGY

parcel	function	payment	credit	services	product

table 3 After this step the similarity, number of associa-

TABLE III
SOME TERMS FOUND IN THE DOMAIN ONTOLOGY

Concept	Depth	Mapped with
payment	1	purl.org/goodrelations/v1#PaymentMethod
method	1	purl.org/goodrelations/v1#PaymentMethod
credit	2	purl.org/goodrelations/v1#PaymentMethodCreditCard
card	2	purl.org/goodrelations/v1#PaymentMethodCreditCard

tions and length is calculated. This leads to calculation of Document Cohesion i.e.0.323164554. The Simplified Dale-Chall is 76.25899281, and the final result for Concept-Based Readability is 0.336277762.

The results of the conceptual based readability using the proposed scheme and Pananya et. al. approach for Konakart, Amazon, click and buy, and ebay WSDLs are show in the figure 2.

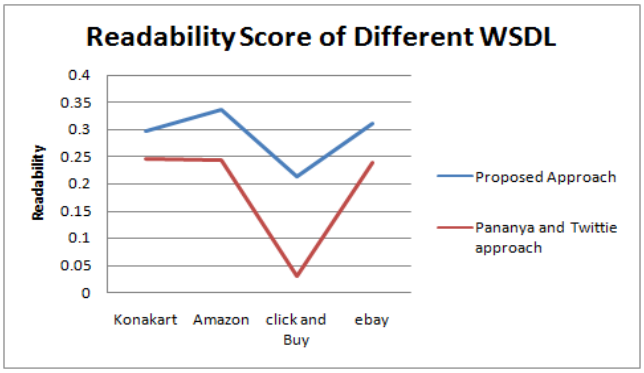

Fig. 2. Readability Score of Different WSDL

The results shows that the Amazon's WSDL is more readable from both approaches, the interesting part of the graph can be seen in ClickandBuy WSDL, where consideration of synonym has significantly increased from the Pananya et. al.approach. So Readability measure increased when unmapped concepts from WSDL are consider in the synonym form. The increased in the resulting values by using the proposed approach shows that some of the words that were not actually matched in the domain ontology were existing in a synonym form. So by using the Pananya et. al.approach we were missing some of the words that may have a match in the domain ontology. The screen-shot of the proposed system is shown in the figure 3.

VI. CONCLUSION

Readability of a Web Service Description helps the consumer in understanding Web Service easily. This paper presents a technique to calculate Conceptual Readability of WSDL documents. If Readability index obtain is high it will

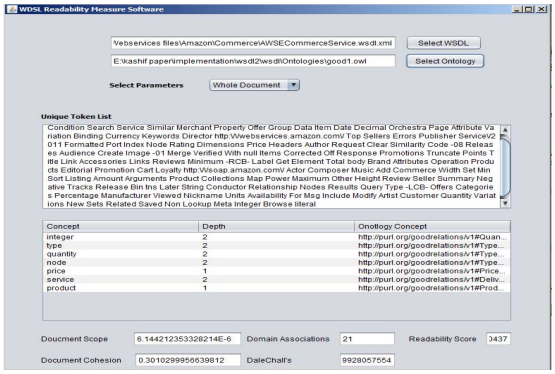

Fig. 3. Conceptual readability calculator Interface

result in high readability and vice versa. To find a more accurate value of Readability index, synonym of unmapped words from the WSDL file are considered, If a word is not found it can be added to ontology so it can be enriched. The proposed approach was tested and evaluated against some WSDL documents. It was shown that readability of WSDL document become higher if missing concepts were considered in calculating it.

REFERENCES

[1] Abid, A.S.; Javed, M.Y., "An Effective Approach for Association Identification in Real Scenarios," Computer and Electrical Engineering, 2009. ICCEE '09. Second International Conference on , vol.1, no., pp.372,374, 28-30 Dec. 2009

[2] World Wide Web Consortium (W3C), Web Services Description Language (WSDL) 1.1. [Online]. Available: http://www.w3.org/TR/WSDL/, Jan. 2014.

[3] Asif Sohail Abid, Atif Sohail Abid, Muhammad Younus Javed, and Adnan Zahid, "Did you Find Any Public Web Service Description over the World Wide Web?," Journal of Advances in Computer Networks vol. 1, no. 2, pp. 143-147, 2013.

[4] Vinoski, S., "Integration with Web services," Internet Computing, IEEE , vol.7, no.6, pp.75,77, Nov.-Dec. 2003.

[5] Qi Yu, Xumin Liu, Athman Bouguettaya, and Brahim Medjahed. 2008. Deploying and managing Web services: issues, solutions, and directions. The VLDB Journal 17, 3 (May 2008), 537-572.

[6] E. A. Masri and Q. H. Mahmoud, Investigating web services on the world wide web, International World Wide Web Conference (WWW 2008), pp. 795-804, 2008.

[7] Pikulski, Readability, Houghton Mifflin Harcourt. [Online]. Available: http://www.eduplace.com/state/author/pikulski.pdf, Jan. 2014.

[8] Sripairojthikoon, P.; Senivongse, T., "Concept-based readability of web services descriptions," Advanced Communication Technology (ICACT), 2013 15th International Conference on , vol., no., pp.853,858, 27-30 Jan. 2013

[9] Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

[10] X. Yan, D. Song, and X. Li, "Concept-based document readability in domain specific information retrieval," Proc. 15th ACM Int. Corif. Information and Knowledge Management (CIKM 2006), 2006, pp. 540-549.

[11] J. S. Chall and E. Dale, Readability Revisited: The New Dale-ChallReadability Formula. Brookline BookslLumen Editions, 1995.

[12] E-Business and Web Science Research Group, GoodRelations Language Reference Version 1.0. [Online]. Available: http://www.heppnetz.de. April. 2014.

[13] http://developer.ebay.com/webservices/latest/shoppingservice.wsdl

[14] https://api.clickandbuy.com/webservices/soap/pay_1_1_0.wsdl