iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** I **Month of publication:** January 2019

DOI: http://doi.org/10.22214/ijraset.2019.1098

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue I, Jan 2019- Available at www.ijraset.com

Development of an Improved Schema Entropy and Interface Complexity Metrics

Sotonwa K. A.[1], Olabiyisi S. O.[2], Omidiora E. O.[3], Oyeleye C. A.[4]
[1]*Department of Computer Science and Information Technology, Bells University of Technology, Ota, Ogun State, Nigeria*
[2,3,4]*Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.*

Abstract: Extensible Markup Language is a mark-up language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable for managing information contained in the schema documents and exchanging wide variety of elements class of data to reflect understandability and reusability. The developed schema entropy and interface complexity metrics are based on the concept from DTD metrics to measure complexity of schema on similarly structured elements, distinct structured elements and their occurrences using the Number of Attributes (NOA), Number of Equivalence Class (NOC), Frequency Occurrence of the class (FOC) and Number of Elements (NOE). A higher value of ISEM and IICM tend to a degree of high flexibility and reusability quality. The empirical validation of the metrics is applied on 40 WSDL schema files. A comparison with similar measures is also performed. The proposed ISEM and IICM metrics validated practically and theoretically. The comparative study proves the robustness of the metrics and performed better in terms of reusability and reducing lengthy code. The statistical analysis of the study showed a significant and linear relationship; and high degree of correlation.

Keywords: XML, WSDL, RNG, ISEM and IICM.

I. INTRODUCTION

The XML is a simple, very flexible text format derived from SGML (ISO 8879) designed to meet the challenges of large-scale electronic publishing playing an important role in the exchange of a wide variety of data on the web and a new technology for web applications. Web services that are based on XML technologies enable integration of diverse IT processes [1][2][3][4]. Due to its flexible nature and ease of performance, XML [5[6][7] serves very well as a ubiquitous, platform-independent data representation and transport format which is accepted in diverse fields. The effective and proper performance of XML in diverse domains requires well-designed of XML schemas which can be written by a series of XML schema language such as Document Type Definition (DTD)[8], World Wide Web Consortium XML Schema (WXS)/XML Schema Definition (XSD)[9][10]11][12]. Schematron and Regular Language for Next Generation (Relax NG)[13][14]. The XML documents play very significant task in software development process and schema metrics can be developed for software products to enable the quantification of the schema sizes, complexity, quality and other properties[5][6] [7][15][16]. The XML documents used in this paper are acquired from Web Service Description Language (WSDL) and implemented in RNG[17][18][19][20[[21[22][23].

II. PROBLEM STATEMENT

In order for XML documents to provide a common understanding about data exchange between applications, XML documents should be validated against the XML Language. Existing metrics consider number of class elements without considering their attributes, since schema documents consist of elements and attributes, hence it was essential to devise improved measures that take note of both class elements and the number of attributes in calculating the structural complexity of schema documents. This paper attempts to evaluate the performance of the metrics with the schema documents acquired in WSDL and implemented in Relax NG. The paper is aimed at formulating metrics to help provide possible solution for flexibility, maintainability and reusability of XML documents.

III. LITERAATURE REVIEW

Out of the numerous proposed measures, selecting a particular complexity measure is a problem, as every measure has its own features and limitations. Although XML based web applications are important, metrics for XML schema document are scarce and there has been very little research to create quality metrics for XML schema documents for improving the web engineering process. Thus, a mature process can produce high quality schema documents. Some of the existing schema metrics are:

Misra and Basci[6] metric measured the assessment of the structural complexity based on schema entropy concept and intended to measure the complexity of the schema documents in XSD. SE made it obvious that understanding the structure and the relation between the nodes of a binary tree is easier than that of an irregular tree thus this provided more information about the understandability and maintainability but the metric failed to reflect the reusability of the schema documents in comparing schemas of equal number of complex type definitions. No theoretical validation was carried. Basci and Misra[7] developed measures that were targeted at finding the structural complexity of DTD schema languages. The entropy metric E(DTD) was adopted from communication information theory and distinct structured element repetition scale DSERS(DTD) was adopted from ARS metric. It was found that measuring the complexity of the XML schema documents are more realistic and can be useful in differentiating DTDs, which have the same size. However, the metrics failed to address the issue of limited possibilities of expressing class element in any order which have different sizes; also these measures have only been applied in the DTD. The metrics were validated theoretically using Weyuker's properties and satisfied six properties. Thaw and Misra[24] formulated Entropy Measure of Complexity metric (EMC) intended to measure the reusable quality of XML schema documents based on the entropy concept, inheritance feature elements and attributes. The drawback of this measure was that no software tool had been implemented to aid the measurement thus making it complex and unlikely to be adopted in the industry. It satisfied eight of the Weyuker's properties. Maja *et al.*[25] defined full set of six composite metrics for measuring each building block / concept properties such as structure, clarity, optimality, minimalism, re-use and flexibility for assessing an XML schema quality. There was restricted access to full standard XML therefore it was difficult to define documentation of XML schema. No theoretical validation was conducted. Falola *et. al.*[26] evaluated and made comparison of metrics for XML schema language which is based on their unique features, advantages and limitations. In addition, the study also discussed whether or not theoretical, practical and empirical validations had been conducted on the various metrics.

IV. METHODOLOGY

A. *The Metrics (ISEM and IICM)*

The metrics are applied on 40 different schema files acquired from WSDL and implemented in Relax NG. The RNG codes were different from each other in their architecture and the calculation of ISEM and IICM for different schema documents are based on graphs structure from the codes and the listings of the graph. The development of ISEM and IICM considered Number of Attributes (NOA) is the number of feature used to describe a property or to provide additional information about an element in a particular schema document, Number of Equivalence Class (NOC) is the number of equivalence class that reflects the number of unique element structures in the schema documents, Frequency Occurrence of the Class (FOC) is the member count of each class which reflects the number of occurrences of each class member, I^{th} Class is the number occurrence of an elements node, Number of Element Nodes (NOE) is the number of elements nodes in a particular schema document, Edges are directed lines that connect two elements together which represents parent-child relationship between the elements of RNG schema in a directed graph and Fanning is the ratio of edges to NOE.

Leveraging on the entropy concept from DTD Metric[7] using communication information theory[27] defined as a measure of uncertainty or variety. Then, the value of entropy can be calculated as:

$$E = -\sum_{i=1}^{n} P(C_i)\log_2 P(C_i) \quad (1)$$

Incorporating *NOA* into the entropy to take care of the account for the number of attributes, entropy described as a given schema documents having *NOC* distinct class of elements can be calculated using relative frequency as unbiased estimated of their probability. Therefore equation (1) is written as:

$$ISEM = -\sum_{i=1}^{NOC} (FOC_i)\log_2 (FOC_i) + NOA \quad (2)$$

Where *NOC* is number of equivalence class, i^{th} class is the number occurrence of an element class, FOC_i is the frequency occurrence of i^{th} class and *NOA* is the number of attributes.

As entropy value of a schema document is calculated by considering classes of the schema documents which was the number of elements and attributes inside the equivalence class divide by the total number of elements and attributes of the schema documents, the Interface complexity metric (DSERS)[7]:

$$DSERS = \sum_{i=1}^{p} de_i^{\,2} / \#e \quad (3)$$

Where *p* is the number of equivalence classes, de_i is the number of members inside the i^{th} class, *#e* is the total number of element nodes in schema documents. Also IICM is formulated by integrating *NOA* as well to take care of the account of number of attributes. The equation (3) is changed as follows:

$$IICM = \sum_{i=1}^{noc} (FOC_i^{\,2} / NOE) + NOA \quad (4)$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue I, Jan 2019- Available at www.ijraset.com

Where n is the number of equivalence class, FOC_i is the frequency occurrence of class, *NOE* is the number of element nodes in the schema documents and *NOA* is the number of attributes an element of RNG has in a particular schema documents

B. Analysis of ISEM and IICM

Samples of the demonstration of the proposed metrics for Saludar implemented in RNG with the directed graph representation is shown in the Figures 1 and 2, the Listing is given in Figure 3 and the analysis are given blow:

```
<?xml version="1.0" encoding="UTF-8"?>
<grammar
  xmlns="http://relaxng.org/ns/structure/1.0"
xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0"
datatypeLibrary="http://www.w3.org/2001/XMLSchema-datatypes">
  <start>
    <element name="Saludar">
         <element name="SaludarResponse">
                 <zeroOrMore>
             <element name="SaludarResult" >
               <data type="string" />
             </element>
         </zeroOrMore>
               </element>
                    <element name="Saludo">
                  <zeroOrMore>
              <element name="nombre" >
                 <data type="string" />
               </element>
                     </zeroOrMore>
                          </element>
                             <element name="SaludoResponse">
                         <zeroOrMore>
                     <element name="SaludoResult" >
                          <data type="string" />
                        </element>
                   </zeroOrMore>
                        </element>
                   </element>
                     </start>
              </grammar>
```

Figure 1: RNG Code for Saludar

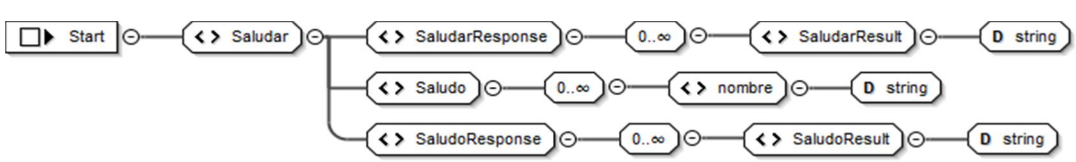

Figure 2: Directed Graph Representation of RNG Schema for Saludar

C_1 = {Saludar}
C_2 = {SaludarResponse, Saludo SaludoResponse}
C_3 = {SaludarResult, nombre, SaludoResult}

Figure 3: Listing for Saludar Schema Files

$$ISEM = -\sum_{i=1}^{3}(FOC_i)\log_2(FOC_i)+NOA$$

= - = $[1/7*\log_2(1/7) + 3/7*\log_2(3/7) + 3/7*\log_2(3/7)] + 0 = 1.4485$

$$IICM = \sum_{i=1}^{3} FOC_i^{\,2} \,/\, NOE + NOA \quad (3)$$

$= 1^2/7 + 3^2/7 + 3^2/7 + 0 = 2.7142$ Fanning = e/n = 6 /7 = 0.8571

V. RESULTS

Series of experiments were conducted to show the effectiveness of the proposed metrics, analyses of all the implemented RNGs can be seen in Table 1.

A. *Empirical Validations of ISEM and IICM*

It was observed that some schemas though some have similar number of NOE yet they had different complexities values were different because schema documents with lower complexities values tend to be dominated by fewer distinct structured elements than those that had higher frequency of occurrences which implied that some schema documents had more diversity in their class elements. At the same time some schemas had equal number of NOE and their complexities values are also equal because the lower and higher complexities values for each metric can be known in terms of elements count and generally, more elements reflect the notion that larger RNG were able to contain more repetition. And some schemas do not have anything in common; both their complexity values and NOEs. This is due to the fact that the ISEM and IICM considered diversity in each class element and their frequencies. The schema documents that exhibited greater variety in class elements with less frequency of occurrences had greater value of complexities than the one that exhibited less variety in class elements with high frequency of occurrences. Finally, some schemas had some elements similar; the cause is that in combination of these RNG, reusable global group name can be defined and giving reference to these global elements group within their various defined schema and listing groups

Table 1: Complexity Measure for the Proposed Metrics and Existing Metrics

S/No	Schemas	NOC	NOE	NOA	Edges	SE	IC	Fanning	ISEM	IICM
1	Subset	4	10	1	9	1.7220	3.4000	0.9000	2.7220	4.4000
2	Ping	8	21	0	20	2.6783	3.7619	0.9523	2.6783	3.7619
3	Saludar	3	7	0	6	1.4485	2.7142	0.8571	1.4485	2.7142
4	Translation	3	5	1	4	1.3710	2.2000	0.8000	2.3710	3.2000
5	ValidateCard	3	6	0	5	1.4589	2.3333	0.8333	1.4589	2.3333
6	Getbible	10	19	0	18	3.0040	2.8947	0.9473	3.0040	2.8947
7	Books	18	34	1	33	3.5699	4.7088	0.9705	4.5699	5.7088
8	AddressBook	3	3	1	2	1.5849	1.0000	0.6666	3.5849	3.0000
9	Authorization	8	13	1	12	2.5069	3.3846	0.9230	3.5069	4.3846
10	Mutants	6	13	0	12	2.3529	2.8461	0.9230	2.3529	2.8461
11	StockHeadlines	5	16	1	15	1.6264	7.0000	0.9375	2.6264	8.0000
12	ConvertTemp	3	6	0	5	1.2516	3.0000	0.8333	1.2516	3.0000
13	Links	4	11	2	10	1.4909	5.0000	0.9090	3.4909	7.0000
14	Phone	3	9	2	8	1.9860	5.6666	0.8888	3.9860	7.6666
15	World	9	17	0	16	3.7368	4.1176	0.9411	3.7368	4.1176
16	Advert	6	13	2	12	0.1411	1.2692	0.9230	2.1411	3.2692
17	GetData	3	7	3	6	1.4485	2.7142	0.8571	4.4485	5.7142
18	AccountExits	7	14	2	13	2.4947	3.0000	0.9285	4.4947	5.0000
19	PowerUnits	4	8	1	7	1.8113	2.5000	0.8750	2.8113	3.5000
20	Table	3	5	2	4	1.5037	1.8000	0.8000	3.5037	3.8000
21	Inventory	3	7	3	6	1.4485	2.7142	0.8571	4.4485	5.7142
22	GasMeter	7	16	0	15	2.4772	3.3750	0.9375	2.4772	3.3750
23	GetTariff	3	7	3	6	1.4485	2.7142	0.8571	4.4485	5.7142
24	Lot	9	17	0	16	2.7769	3.2352	0.9411	2.7769	3.2352
25	BonPlan	6	13	2	12	2.1411	3.7692	0.9230	4.1411	5.7692
26	LinearAds	6	12	1	11	2.2208	3.1666	0.9166	3.2208	4.1666
27	Variables	6	12	1	12	2.1252	2.6666	0.9166	3.1252	3.6666
28	Log	8	16	1	15	2.6039	3.3750	0.9375	3.6039	4.3750
29	Bank	2	3	2	2	0.9182	1.6666	0.9375	2.9182	3.6666
30	BIZServices	8	11	2	10	2.0048	3.1818	0.6666	4.0048	5.1818
31	Briefs	10	18	2	17	2.5830	5.0000	0.9444	4.5830	7.0000
32	CalServices	9	15	2	14	2.5947	3.8000	0.9333	4.5947	5.8000
33	Soap	2	5	1	4	0.7219	1.5454	0.8000	1.7219	2.5454
34	Contact	19	43	2	42	2.9535	12.0232	0.9767	4.9535	14.0232
35	ArendsogServices	9	15	2	14	1.8644	11.8800	0.9600	3.8644	13.8800
36	Account	16	35	0	34	2.6542	11.8571	0.9714	2.6542	11.8571
37	Collection	8	14	2	13	2.4031	4.0000	0.9285	4.4031	6.0000
38	VerifyRecord	7	19	0	18	2.2368	2.8461	0.9193	2.2368	2.8461

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue I, Jan 2019- Available at www.ijraset.com

39	EmaiStmp	8	16	0	15	2.7266	4.0000	0.9375	2.7266	4.0000
40	FedACHcities	12	24	1	23	3.3219	2.9166	0.9375	4.3219	3.9166

B. *Theoretical Validation of ISEM and IICM by Weyuker's Properties*

1) *Property 1: (∃P)(∃Q)(|P| ≠ |Q|):* Where *P* and *Q* are program body. This property states that a measure should not rank all programs as equally complex. This property is satisfied since these schemas have different complexities measures and their values were not equal.
2) *Property 2:* Let *c* be a non-negative number then there are only a finite number of programs having complexity *c*. Since schema documents consist of only finite number of elements and attributes, calculation of ISEM and IICM depend largely on the graph structure obtained from schemas and it is also possible to calculate with different non negative values assign to them, then this property is also satisfied by ISEM and IICM metrics.
3) *Property 3:* There are distinct programs P and Q such that |P|=|Q|: this property states that a valid complexity measure allows different schemas to have the same complexity values. One can find the same ISEM and IICM values, if different schema documents have the same fan-in and fan-out, same listing and same graph structure. For example two $RNG_{(GetDsta)}$ and $RNG_{(Inventoy)}$ have equal value for ISE (4.4485) and IIC = (5.7142) as seen from Table 1. Thus, the two different schemas had the same complexity values; so the proposed measures satisfied property.
4) *Property 4: (∃P)(∃Q)(P ≡ Q &|P| ≠ |Q|):* This property states that even though two programs compute the same function, their program complexities depend upon the implementation details. Since all these schemas depend on the element and attribute structures of the schema documents then the property is satisfied by ISEM and IICM.
5) *Property 5:* (∀P)(∀Q) (|P| ≤ |P;Q| and |Q| ≤ |P;Q|): This property states that if the combined program is constructed from programs P and Q, the value of the program complexity for the combined program is larger than the value of the program complexity for P or Q as seen from Figure 3..

$ISEM$ = $RNG_{(Ping:AuthorizationList)} = -\sum_{i=1}^{14}(FOC_i)*\log_2*(FOC_i)+NOA$ = - [(1/31)*$\log_2$*(1/31) + (1/31)*$\log_2$*(1/31) + (9/31)*$\log_2$*(9/31) + (1/31)*$\log_2$*(1/31) + (1/31)*$\log_2$*(1/31) + (1/31)*$\log_2$*(1/31) + (1/31)*$\log_2$*(1/31) + (1/31)*$\log_2$*(1/31) + (1/31)*$\log_2$*(1/31 + (1/31)*$\log_2$*(1/31) + (3/31)*$\log_2$*(3/31) + (5/31)*$\log_2$*(5/31) + (4/31)*$\log_2$*(4/31) + (1/31)*$\log_2$*(1/31)] + 0 + 1 = 3.5705

$IICM$ = $RNG_{(Ping:AuthorizationList)} = \sum_{i=1}^{14}(FOC_i^{\,2}/NOE)+NOA = 1^2+1^2+9^2+1^2+1^2+1^2+1^2+1^2+1^2+1^2+3^2+5^2+4^2+1^2/31+0+1 = 5.5483$

Let P represents the ISEM and Q to represent IICM for Ping, authorization. The complexities values of ISEM for $RNG_{(Ping)}$ = 2.6783 < $RNG_{(Ping.Authorization)}$ = 3.5705, $RNG_{(Authorization)}$ = 3.5069 < $RNG_{(Ping.Authorization)}$ = 3.5705 for IICM with complexities values $RNG_{(Ping)}$ = 3.7619 < $RNG_{(Ping:Authorization)}$ = 5.5483 and $RNG_{(Authorization)}$ = 4.3846 < $RNG_{(Ping:Authorization)}$ = 5.5483. Since the proposed metrics for Ping and Authorization are less than their combination then this property is satisfied by the proposed metrics.

C_1= {Ping}
C_2 = {AuthorizationList}
C_3 = {Login, LoginResponse, GetAuthenticationToken, GetAuthenticationTokenResponse, GetAuthorizationList, PICredentials, RequestAccess, GetAuthorizationListResponse, RequestAccessResponse}
C_4 = {Ping}
C_5 = {localID}
C_6 = {GetAuthorizationListResult}
C_7 = {localID}
C_8 = {RequestAccessResult}
C_9 = {PingResponse}
C_{10} = {PingResult}
C_{11} = {Email, Password, ReturnBaseUrl}
C_{12} = {LoginResult, Success, ErrorCode, AuthenticationMessage, AccountID}
C_{13} = {Email, Password, AccountID, GoToEvelopeID}
C_{14} = {GetAuthenticationTokenResult}

Figure 4: Listing Combination of Ping and Authorization for Weyuker's Property 5

6) *Property 6 (a):* $(\exists P)(\exists Q)(\exists R)(|P| = |Q|)$ & $(|P;R| \neq |Q;R|)$ 6 *(b)* and $(\exists P)(\exists Q)(\exists R)(|P| = |Q|)$ & $(|R;P| \neq |R:Q|)$: states that it is possible to find two RNGs of equal ISEM and IICM values which when separately concatenated to a third RNG yield different values. Let P represents $RNG_{(GetData)}$ with listing C_1 = 1, C_2 = 3, C_3 = 3, Q $RNG_{(Gettariff)}$ with C_4 = 1, C_5 = 3, C_6 = 3 and R $RNG_{(Inventory)}$ with C_7 = 1, C_8 = 3, C_9 = 3 schemas respectively. P and Q had element different in structure and their graph representations were also different to each other but due to similar number of elements and similar number of listings their complexity values were equal. While R had same structure as P and all the 3 RNGs (P, Q and R) did not have any element in common. The complexity values of ISEM for $RNG_{(P)}$ = $RNG_{(Q)}$ = $RNG_{(R)}$ = 4.4485 and IICM for $RNG_{(P)}$ = $RNG_{(Q)}$ = $RNG_{(R)}$ = 5.7142. In combining P and R, the combined RNGs had 3 classes C_1 = 2, C_2 = 6, C_3 = 6 because these two RNGs had same graph representation with no common element. So the complexity values of ISEM for $RNG_{(P:R)}$ = 4.4485 and IICM for $RNG_{(P:R)}$ = 8.4285 as shown in Figure 5. Combining Q and R, resulting classes would be doubled C_1 = 1, C_2 = 3, C_3 = 3, C_4 = 1, C_5 = 3, C_6 = 3 because these RNGs also have fully different elements. The complexity values of ISEM for $RNG_{(Q:R)}$ = 8.4485 and IICM for $RNG_{(Q:R)}$ = 5.7142. See the analysis of Listing in Figure 6. The result show that for ISEM $RNG_{(P:R)} \neq RNG_{(Q:R)}$ and for IICM $RNG_{(P:R)} \neq RNG_{(Q:R)}$ so property 6 a and b is also satisfied by proposed metrics.

C_1 = {Data, Inventory}
C_2 = {GetData, GetDataResponse, ArrayColumnDefinitions, Account, GetDataResult, ColumnDefinition}
C_3 = {AddMeterToInventory, AddMeterToInventoryResponse, DeployElectricMeter Meters, AddMeterToInventoryResult, Meters}

Figure 5: Listing Combination of GetData and Inventory for Weyuker's Properties 6a

ISEM = RNG$_{(P:R)}$ $= -\sum_{i=1}^{3}(FOC_i)\log_2(FOC_i) + NOA$ = - [2/14*log$_2$(2/14) + 6/14*log$_2$(6/14) + 6/14*log$_2$(6/14)] + 3 = 4.4485

IICM= RNG$_{(P:R)}$ $= \sum_{i=1}^{3}(FOC_i^2 / NOE) + NOA$ = $2^2 + +6^2 + 6^2/14 + 3 = 8.4285$

C_1 = {GetTariff}
C_2 = {Inventory}
C_3 = {tariff, GetTariffResponse, HideBanner}
C_4 = {AddMeterToInventory, AddMeterToInventoryResponse, DeployElectricMeter}
C_5 = {GetTariffResult, HideBannerResult, HideBannerResult}
C_6 = {Meters, AddMeterToInventoryResult, Meters}

Figure 6: Listing Combination of GetTariff and Inventory for Weyuker's Properties 6b

ISEM = RNG$_{(Q:R)}$ $= -\sum_{i=1}^{6}(FOC_i)\log_2(FOC_i) + NOA$ = - [1/14*log$_2$(1/14) + 3/14*log$_2$(3/14) + 3/14*log$_2$(3/14) + 1/14*log$_2$(1/14) +3/14*log$_2$(3/14) + 3/14*log$_2$(3/14)] + 3 = 8.4485

IICM = RNG$_{(Q:R)}$ = $\sum_{i=1}^{6}(FOC_i^2 / NOE) + NOA$ = $1^2 + 3^2 + 3^2 + 1^2 + 3^2 + 3^2 /14 + 3 = 5.7142$

7) *Property 7:* There are program bodies P and Q such that Q is formed by permuting the order of the statement of P and (|*P*|≠|*Q*|): This property states that permutation of elements within the item measured can change the metric values. The intent is to measure the metric value due to permutation of schemas. If the place of definitions of elements and attributes were to be changed in RNG documents, the complexities values would also changed because in RNG it is possible to define any local element or attribute as global and make them reusable components. Therefore the property holds for the proposed measures.
8) *Property 8:* If P is renaming of Q, then |*P*| = *Q*|*:* This property requires that when the name of the schema changed it would not affect the complexity value. The values of ISEM and IICM were real numbers so if the name of the schema changed it cannot change the values of the metrics. Hence, the metrics clearly adhere to this property.
9) *Property 9:* (∃*P*)(∃*Q*)(|*P*| + |*Q*|) <(|*P*;*Q*|). This property states that the program complexity of a new program combined from two programs is greater than the sum of two individual program complexities. Using the same example given in Property 5, since ISEM for RNG$_{(Ping:Authorization)}$ < RNG$_{(Ping)}$ + RNG$_{(Authorization)}$ i.e. 3.5705 < 6.1852 Also IICM for RNG$_{(Ping:Authorization)}$ < RNG$_{(Ping)}$ + RNG$_{(Authorization)}$ i.e. 5.5483 < 8.1465, therefore proposed metrics do not satisfy this property. In this section, the proposed metrics are validated against 9 Weyuker's properties. The ISEM and IICM satisfy 8 properties. It is important to note that it is not necessary to satisfy all the Weyuker's properties. From this point of view ISEM and IICM satisfying eight Weyuker's properties shows robust measures.

C. *Comparative Study of ISEM and IICM*

The relative graphs in Figure 6 and 7 showed the comparison among Fanning, proposed metrics (ISEM and IICM) and existing measures (SE and IC). Close inspections of these graphs showed that NOA were closely related and considered in the proposed metrics rather than in existing metrics, thus this differentiate their complexities values and implied that schema documents that had many inheritance features of elements would give greater complexity values because of high degree of flexibility and reusability quality. The higher fanning values for RNG could be interpreted as that elements were highly connected and dependent to each other. Thus, any modification made in any individual element would update the other element to which that individual element was connected

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue I, Jan 2019- Available at www.ijraset.com

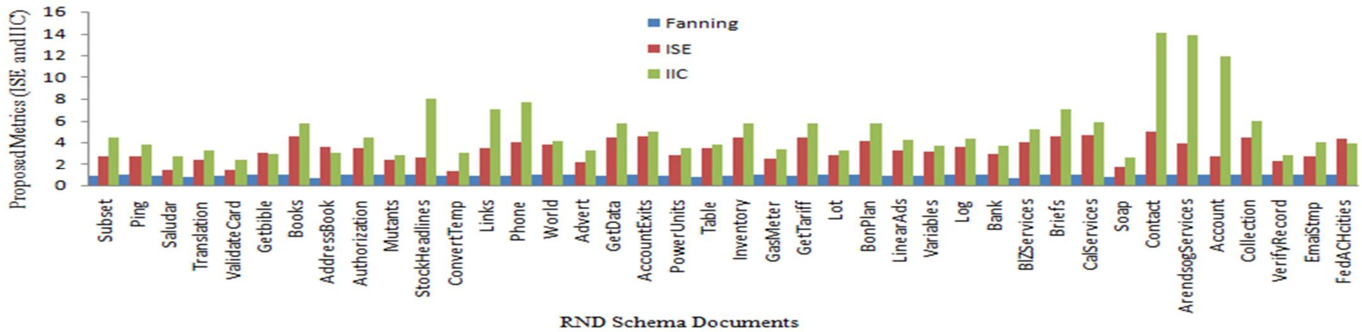

Figure 7: Relative Graph between Fanning and ISEM and IICM

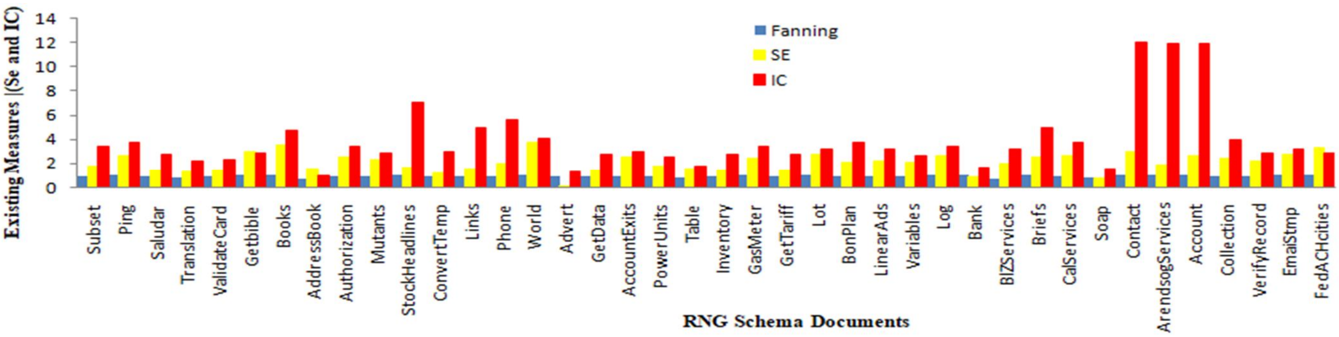

Figure 8: Relative Graph between Fanning and Existing Metrics (SE and IC)

These graphs depicted in figure 8 and 9 reflect the comparison results between the ISE and IIC metrics and existing metrics. There was inverse relation between ISE and IIC because the lower the ISE the higher IIC values; thus, this had the same meaning which is lower psychological complexity of RNG, so values of these metrics were not in contradiction. However, Figure 7 showed lower complexity value to the ISE and IIC metrics hence, this exhibited more regularity and understanding because it provided more information about the schemas.

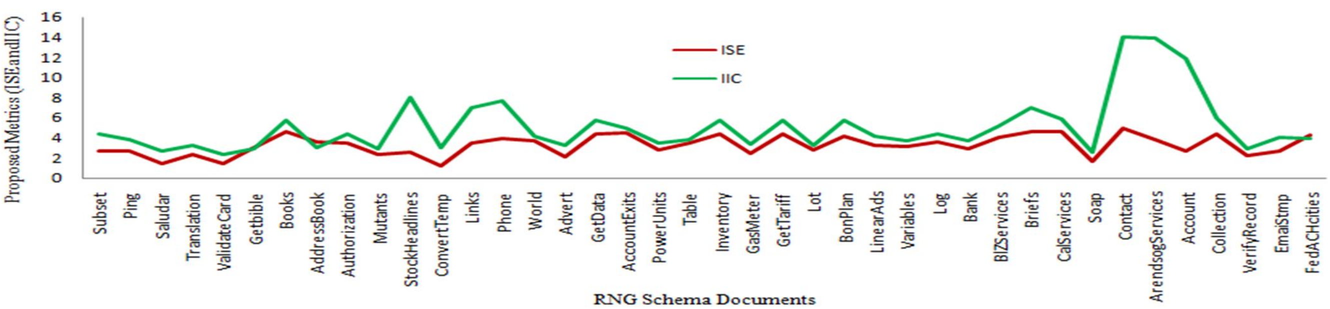

Figure 9: Comparison Result between ISEM and IICM

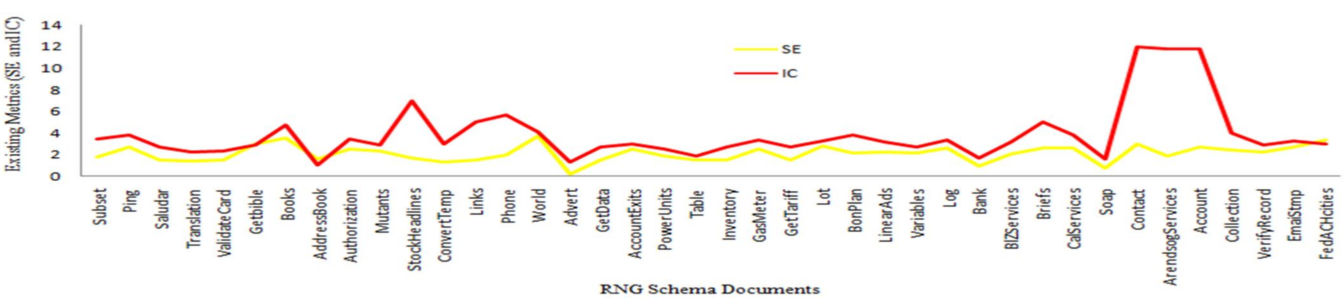

Figure 10: Comparison Result between SE and IC Metrics

D. Degree of Investigation using Pearson Correlation coefficient and Analysis of Variance (ANOVA)

Table 2 showed the Pearson correlation coefficient for, fanning, ISEM and IICM measures, the Pearson correlation coefficient value of 0.528 indicated a positive moderate relationship between fanning and ISEM and this yielded the computed probability values (p) of 0.000 < 0.05 level of significant therefore this showed that a correlation exists between fanning and ISE so, P-values rejects H_0. The correlation value of 0.560 showed a positive moderate relationship between fanning and IICM and the 2-tailed significant value of 0.00 < 0.05 designated a significant correlation between fanning and IICM so H_0 rejects P-values. Also the Pearson correlation value of 0.783 indicated a positive strong relationship between ISEM and IICM, this also yielded P-value of 0.00 < 0.05 level of significant so there was a significant correlation between ISEM and IICM as a result of this P-value rejects H_0.

Table 2: Pearson Correlation of Complexity Values for Different Measures of ISEM and IICM in Relax NG

		Fanning	ISEM	IICM
Fanning	Pearson Correlation	1	.528**	.560**
	Sig. (2-tailed)		.000	.000
	N	40	40	40
ISEM	Pearson Correlation	.528**	1	.783**
	Sig. (2-tailed)	.000		.000
	N	40	40	40
IICM	Pearson Correlation	.560**	.783**	1
	Sig. (2-tailed)	.000	.000	
	N	40	40	40

Table 3 showed the correlation table for Fanning, SE and IC measures, the Pearson correlation value of 0.582 indicated a positive moderate relationship between Fanning and SE. This yielded P values of 0.00 < 0.05 level of significant hence, there is a correlation existence between Fanning and SE thus, P-value rejects H_0. The correlation value of 0.527 showed a positive moderate relationship between Fanning and IC, the 2-tailed significant value of 0.00 < 0.05designated a significant correlation between Fanning and IC so H_0 is rejected by P-values. Correlation value of 0.182 implied a negative weak relationship between SE and IC, which showed that the 2-tailed significant value of 0.261 > 0.05 therefore a non-significant correlation exist between SE and IC thus, P-value accepts $H_{0.}$

Table 3: Pearson Correlation of Complexity Values for Different Measures of Existing Metrics in Relax NG

		Fanning	SE	IC
Fanning	Pearson Correlation	1	.582**	.527**
	Sig. (2-tailed)		.000	.000
	N	40	40	40
SE	Pearson Correlation	.582**	1	.182
	Sig. (2-tailed)	.000		.261
	N	40	40	40
IC	Pearson Correlation	.527**	.182	1
	Sig. (2-tailed)	.000	.261	
	N	40	40	40

Table 4 is the ANOVA for fanning, ISEM and IICM and it was obvious that the model is significant considering both the F-Statistics (75.4794) greater than the F-Tabulated value of $F_{2.117}$ (3.15) degrees of freedom and the significant value (0.000) is less than alpha value (0.05). Since 0.000 < 0.05 for the proposed metrics, the null hypothesis H_0 is rejected, therefore the outcome show that there is significant relationship between fanning, ISE and IIC metrics. . It was discovered that there exist linear relationship and high degree of correlation between the different measures of ISEM, IICM, SE and IC in RNG.

Table 4: ANOVA for ISEM and IICM

Sources of Error	Sum of Squares	Variance Estimate	DF	F
Between Groups	SS_b (351.7175)	S_b^2 (175.8737)	3	53.2869
Within Groups	SS_w (386.1607)	S_w^2 (3.3005)	117	$F_{0.01, 2, 117} = 3.15$

Total	737.8782	179.1742	120

VI. CONCLUSION

The ISEM and IICM make more sensitive measurement, so information contained in the WSDL was provided and also measure the difficulties in understanding the schemas. SE and IC were not able to measure class elements comprehension, of a fact empirical validation have shown that ISEM and IICM were able to reflect strong support for class elements to make them appear in any order. SE and IC based on total number of occurrence s of input and output parameters, counting the number of input and output is not clear and ambiguously interpreted. Where, ISEM and IICM were able to handle those issues. ISEM and IICM metrics implemented in RNG schema is highly structured and can partner with other schema language with a separate data typing language which makes it simpler in exhibit a better presentation of a given schema document than SE and IC. Future research may be geared towards evaluating the cognitive complexity of the XML schema documents for XSD, DTD and Relax-NG and also since XML has been used by databases, another metric can develop a criterion to evaluate and maintain the quality of the XML enabled database in future.

RETERENCES

[1] Cerami E. (2002): Web Services Essentials Distributed Application with XML-RPC, SOAP, UDDI & WSDL, O'Relly Publisher.
[2] Erl T. (2004): Service Oriented Architecture: A Field Guide to Integrating XML and Web Services, Prentice Hall Publishers.
[3] Newcomer E. and Lomow G. (2004): Understanding SOA with Web Services, Addition Wesley Professional
[4] Thomas E. (2004): Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services, Prentice Hall Publishers, 2004.
[5] Basci D. and Misra S. (2009): Measuring and Evaluating Design Complexity Metric for XML Schema Document, Computer Journal of Information Science and Engineering, 25th September 2009, vol. 25, no. 5, pp. 1405-1425.
[6] Basci D. and Misra S. (2010): Entropy as a Measure of Quality of XML Schema Document, The International Arab Journal of Information Technology, 8 (2011), 75-83.
[7] Basci D., Misra S. (2011): Document Type Definition (DTD) Metrics, Romanian Journal of Information Science and Technology Vol. 14 no 1, pp 31-50.
[8] Bray T., Jean P. and Sperberg-McQueen M.C. (2004): Extensible Markup Language (XML) 1.0 W3C Recommendation, February 1998, World Wide Web Consortium (W3c) URL: http//www.w3.org/TR/1998/REC-xml-19980210.html.2012-02-02.
[9] Binstock C., Peterson D., Smith M., Wooding M., Dix C. and Galtenberg C. (2002): The XML Schema Complete Reference, Addison Wesley Professional Publishers.
[10] Thompson H. S, Beech D., Muzmo. M. and Mendel- sohn N. (2004): XML Schema Part 1: Structures Second Edition, W3C Recommendation, 2004. http://www.w3.org/TR/xmlschema-1/
[11] David C. F. (2001): XML Schema Part 0: Primer" [Online] 30 March 2001 (2001-03-30), W3C, XP002305611 Retrieved from the Internet: URL:http://www.w3.org/TR/2001/PR-xmlschema-0-20010330/.
[12] Lee D. and Chu W. (2000): Comparative Analysis of Six XML Schema Language, ACM SIGMOD Record, 29, 3, pp. 1 {12.
[13] Makoto M. (2002): Relax (Regular Language Description for XML). Retrieved June 2010 from http://www.xml.gr.jp/relax/
[14] ISO: (2002). ISO/IEC TR 22250-1:2002 - Information Technology -- Document Description and Processing Languages -- Regular Language Description for XML (RELAX) -- Part 1: RELAX Core". ISO. Retrieved 2009-12-26.
[15] Basci D., Misra S. (2011a): Metrics Suite for Maintainability of eXtensible Markup Language Web Services, Received 21st January 2010, Revised on 19th November, 2010, doi:10.1049/iet-sen.2010-89, ISSN 1751-8806, www.ietdl.org, IET Software 2011, Vol. 5, Issue 3, pp 320-341
[16] Misra S. (2011b) Evaluation Criteria for Object-oriented Metric, Acta Polytechnica Hungarica, 8(5) pp. 109-136
[17] http://docbook.sourceforge.net/release/dsssl/current/dtds/
[18] http:// java.sun.com/dtd/
[19] http://struts.apache.org/dtds/
[20] http://jonas.objectweb.org/dtds/
[21] https://www.researchgate.net/publication/319877377
http://www.ncbi.nlm.nih.gov/dtd/
http://www.cs.helsinki.fi/group/doremi/publications/XMLSCA2000.html
http://www.pramati.com/dtd/
http://www.w3.org/TR/REC-xml-names/
http://www.omegahat.org/XML/DTDs/
http://www.openmobilealliance.org/Technical/dtd.aspx
http://www.python.org/topics/xml/dtds/
http://www.okiproject.org/polyphony/docs/raw/dtds/
http://www.w3.org/TR/wsdl, Last Visited 2008.
http://www.w3.org/XML/, Last Visited 2008.
http://www.xml.gr.jp/relax, Last Visited 2008.
http://www.w3. org/TR /2004/REC- xmlschema- 1-20041028/, Last Visited 2008.

http://www.w3. org/TR /2001/PR- xmlschema-0-20010330/, Last Visited 2008.
http://www.w3. org/TR /2004/REC- xmlschema-2-20041028/, Last Visited 2008.
http://www.w3.org/TR/1998/REC-xml-19980210, Last Visited 2008.
http://www.xfront.com/GlobalVersusLocal. html, Last Visited 2008.
http://www.oreillynet.com/xml/blog/2006/05/metrics_for_xml_projects_1_ele.html, Last Visited 2008.
[22] http:// ivs. cs. uni- magdeburg. de/sw- eng/ us/ metclas /index.shtml, Last Visited 2008.
[23] http://fisheye5.cenqua.com/browse/glassfish/update-center/dtds/
[24] Thaw T. and Misra S. (2013): Measuring the Reusable Quality for XML Schema Documents Acta Polytechnic Hungarica Vol. 10, No. 4, 2013 pp 87-106.
[25] Maja P., Marjan H., Zoran B. and Bostjan S (2014): XML Schema Metrics for Quality Evaluation, Computer Science and Information Systems 11(4):1271–1289 pp DOI:10.2298/CSIS140815077P Faculty of Electrical Engineering and Computer Science, Smetanovaulica 17, 2000 Maribor, Slovenia and Faculty of Sciences, University of Novi Sad, Trg D. Obradovića 4, 21000 Novi Sad, Serbia.
[26] Falola O., Misra S., Adewumi A. and Damasevicius R. (2017): Evaluation and Comparison of Metrics for XML Schema Languages. Article in Frontiers in Artificial Intelligence and Applications June 2017 DOI: 10.3233/978-1-61499-773-3-
[27] Shannon C. and Weaver W. (1949): The Mathematical Theory of Communication, University of Illinois Press, Urbana.