Implementation of an Improved Interface Complexity Metric on Regular Language for Next Generation

Chistopher A. Oyeleye
Information Systems
Ladoke Akintola University of Technology, Ogbomoso.
caoyeley@lautech.edu.ng

Kehinde A. Sotonwa
Computer Science/Cyber Security
Caleb University Imota
Lagos State, Nigeria
adebola.sotonwa@calebuniversity.edu.ng

Elijah O. Omidiora
Computer Engineering
Ladoke Akintola University of Technology, Ogbomoso
Oyo State, Nigeria
eoomidiora@lautech.edu.ng

Stephen O. Olabiyisi
Computer Science
Ladoke Akintola University of Technology, Ogbomoso
soolabiyisi@lautech.edu.ng

Emmanuel Abiodun
Computer Science
Kwara State Polytechnic, Ilorin,
etabiodun1492@gmail.com

Bello O. Alabi
Mathematics/Physics Science
Afe Babalola University, Ado Ekiti
bellooa@abuad.edu.ng

ABSTRACT
User Interface design metric assist developers to evaluate interface designs in early phase before delivering the software to end users. Controlling and minimizing software complexity is one of the most important objectives of each software development paradigm because it affects all other software quality attributes like reusability, reliability, testability, maintainability etc. This paper presents Improved Interface Complexity (IIC) Metric using Number of Equivalence Class (NEC), Frequency Occurrence of Class (FOC_i), Number of Elements (NE) of the schema documents, the Number of Attributes (NA) and Element Fanning (EF) of an RNG. The proposed metric was applied on real schemas documents data acquired from Web Service Description Language (WSDL) and implemented in Regular Language for Next Generation (RNG). The result showed that RNG reduce complexity of class elements, showed more reusability and flexibility traits and overall understanding of the schema documents becomes much easier which reduces maintenance effort.

Key words: Schema Documents, RNG, WSDL, Improved Interface Complexity, XML Schema Language.

1. INTRODUCTION

The eXtensible Markup Language (XML) is playing important role in the exchange of variety of data on the web. XML as a new technology required well design of XML schema; it is formed as tree structure contains root (parent element) and branches (child /sub-child elements) (Jennifer, 2006). In developing web application some qualities such as reusable, flexible and maintainable (Maja *et al*,. 2014) must be considered to determine the complexity of application software in terms of interaction with digital environment (Thaw and Khin, 2013; Sankar and Irudhyaraj, 2014). The design of interfaces must be assessed precisely in order to control the impact of any required change (Romano and Pinzger, 2011; Dig and Johnson, 2005).To help maintainers improve the software quality, there has been recently an important progress in the area of software automatic refactoring and optimization of code quality (Mens and Tourwe, 2004; Misra and Cafer, 2011; Sotonwa *et al.,* 2014).

To ensure proper data exchange between applications XML documents must be validated against the XML schema language. XML schema language is a description of a type of XML document, typically expressed in terms of constraints on the structure and content of documents of that type, above and beyond the basic syntactical constraints imposed by XML itself (Makoto *et al.,* 2001). These constraints are generally expressed using some combination of grammatical rules governing the order of elements. Though there are numbers of schema languages available (Sotonwa *et al.,* 2019a; Sotonwa, 2020), this research is focusing on Regular Language for Next Generation (RNG). Each language has its own advantages and disadvantages (Marconi and Nentwich, 2004).

RNG is an easy-to-learn schema language that possesses both XML syntax and compact non-XML syntax (Clark and Makoto, 2001). This language can specify patterns for the structure and content of an XML document in a relatively simple but powerful way. It allows attributes to be treated as elements in content models (Dongwon and Wesley, 2000). Most RNG schemas can be algorithmically converted into other schema languages such as XML Schema Definition (XSD) and Document Type Definition (DTD). Nevertheless, RNG has no ability to apply default attribute data to an element's list of attributes (Makoto *et al.*, 2005). RNG has an equivalent form that is much more like a DTD, but with greater specifying power known as the compact syntax. Tools can easily convert between these forms with no loss of features or even commenting (DeRose, 1997). RNG provides very strong support for unordered content which allow sequence of patterns to appear in any order, it also allows for non-deterministic content models i.e. it allows the specification of a sequence, it allows attributes to be treated as elements in content models, (Sotonwa *et al.*, 2019b)

2. RELATED WORK

Misra and Basci (2010) metric measured the assessment of the structural complexity based on Schema Entropy (SE) concept. 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. No theoretical validation was carried. Basci and Misra (2011) developed measures E(DTD) and DSERS(DTD) that were targeted at finding the structural complexity of DTD schema language. These metrics exhibited a better representation of structural complexity of a given schema document. It was found E(DTD) and DSERS(DTD) metrics were more realistic and could be useful in differentiating DTDs, of the same sizes. However, the metrics failed to address the issue of limited possibilities of expressing class element in any order which have different sizes and the measures is only applied on DTD, other schema language could also be used to see the effectiveness of the metrics. The metrics were validated theoretically using Weyuker's properties and satisfied six properties.

Maja *et al.* (2014) defined full set of six composite metrics for measuring each building block of 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. Falola *et al.* (2017) evaluated and made comparison of metrics for XML schema language 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.

Sotonwa *et al.* (2019a) outlined the three common XML languages (DTD, XSD and RNG) as applied to schema metrics and figured out different types of validation process done on each metric which show how recent, effective and comprehensive measure they are. Sotonwa *et al.* (2019b) measured SLOC metric for RNG schema documents to predict the amount of effort required to develop a program. Different types of SLOC were considered on forty (40) different schema files that gave details of all line of codes; whether helping in the efficiency of code execution or not. The metric was used to estimate schemas productivity, maintainability and reliability. But theoretical validation was not conducted on the metric.

From the review of the related works, it was discovered that the existing (DSERS) measures thought of class elements and their number of occurrences but did not consider attributes. Attribute helps to distinguished one schema from the other based on the response time it takes to validate codes, more so, schema documents consist of elements and attributes. Hence, the measure can be improved upon by taking into account the number of attributes; then replaced class elements and number of occurrence with frequency of occurrence, number of equivalence class also replaced with number of elements. Therefore, reflect the difficulty of XML code understandability.

3. RESEARCH METHODOLOGY

3.1 The Approach

The development of IIC metric for XML schema document is applied on forty schemas; these schemas were acquired online from web service description language, and then implemented in RNG. The following approaches were adopted:

(i) The factors affecting the complexity of the schema documents were analyzed by calculating the number of attributes (NA) which 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 (NEC) reflects the number of unique element structures in the schema document, frequency occurrence of the class (FOC_i) is the member count of each class which reflects the number of occurrences of each class member, the number of elements (NE) is the total number of elements in a particular schema document and the Element Fanning (EF) is the ratio of edges to NE.
(ii) Formulate metric to measure IIC metric using the analyzed factors affecting the complexity of schema documents such as NA, NEC, FOC, NE and EF.
(iii) Implementation of the IIC metric on analyzed factors.
(iv) The metric is empirically validated using forty (40) real schemas documents acquired online from WSDL and implemented in RNG. The metric is applied to determine the efforts required in understanding the information contents of each of the implemented schema document.
(v) Then results were compared with the existing metric to prove its improvement.

3.2 Improved Interface Complexity (IIC) Metric

Improved Interface complexity metric is adopted from Argument Repetition Scale (ARS) metric proposed by Boxall and Araban (2004) as an interface metric for measuring the consistency of the arguments that software interface uses (Basci and Misra, 2011). It can be used to measure the interface complexity of the class of a schema documents defined as:

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

Where p is the number of equivalence classes
de_i is the number of members inside the i^{th} class and
$\#e$ is the total number of element nodes in the graph.

Now, the IIC metric was formulated by incorporating NA as well to take care of the account of number of attributes and replacing de_i with FOC_i, #e is changed to NE and p also changed to NEC, hence, Equation (1) is re-written as:

$$IIC = \sum_{i=1}^{NEC} \left(FOC_i^{\,2} / NE \right) + NA \qquad (2)$$

Where NEC is the number of equivalence classes
FOC_i is the frequency occurrence of the class
NE is the number of element in the schema document and.
NA is the number of attributes an element of RNG has in a particular schema document.

3.3 Element Fanning (EF) Metric

Element Fanning metric based on Information Flow metric measured complexity as a function of fan in and fan out (Henry and Kafura, 1981). Fan-in of a procedure is defined as the number of local flows into that procedure plus the number of data structures from which that procedure retrieves information while Fan-out is defined as the number of local flows out of that procedure plus the number of data structures that the procedure updates. Fan-in and fan-out measurement can be defined for files, procedure and object. Thus, a module complexity can be defined as:

$$C = [procedurelength * [fan-in * fan-out]^2 \qquad (3)$$

To better understand and visualize it as graph by taking module (file, procedure or object) as node (which represents the number of element) and call between them as edges (which represents directed lines that connect two elements together i.e. parent-child relationship between the elements of RNG schema in a directed graph). This can be interpreted that as the fanning value increases so does the complexity of a given schema document, therefore equation (3) is also re-written as:

$$EF = \frac{e}{NE} \tag{4}$$

Where *e* is the edges in the schema document and
NE is the nodes in the schema document that represent number of element

3.4 Empirical Validation of IIC Metric

Demonstration of IIC metric is given as a sample of schema documents: saludar.rng and translation.rng. The directed graph representations were given and the empirical validations of the equivalence classes were also given in listings forms from Figures 1-4, respectively. The analyses of the resulting listing were calculated from Equations 2 and 3 respectively:

3.4.1 Analysis of RNG schema documents for Saludar

```
<?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>
```

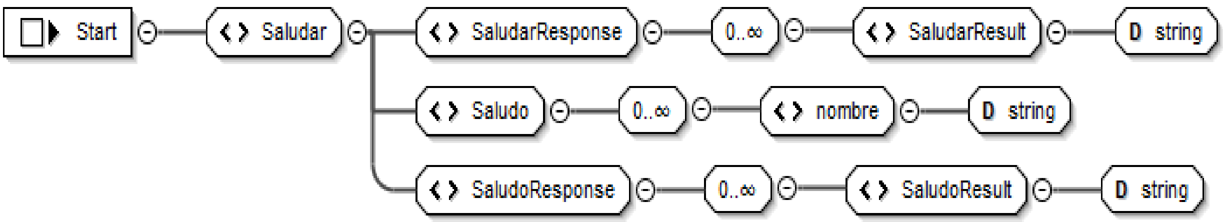

Fig 1: Directed Graph Representation of Schema Document saludar.rng

C_1= Saludar

C_2 = SaludarResponse, Saludo, SaludoResponse

C_3 = SaludarResult, nombre, SaludoResult

Fig 2: Listing of Equivalence Classes of saludar.rng

(i) $$IIC = \sum_{i=1}^{3}\left(FOC_i^{\,2} / NE\right) + NA$$

$$= \left(\frac{1^2+3^2+3^2}{7}\right) + 0 = 2.7142$$

(ii) $$EF = \frac{e}{NE} = \frac{6}{7} = 0.8571$$

3.4.2 Analysis of RNG schema documents for Translation

```
<?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="Translate">
      <element name="LanguageMode">
        <ref name="Language"></ref>
      </element>
      <zeroOrMore>
        <element name="text">
          <data type="string"></data>
        </element>
      </zeroOrMore>
            <element name="TranslateResponse">
              <zeroOrMore>
          <element name="TranslateResult" >
            <data type="string" />
          </element>
        </zeroOrMore>
            </element>
    </element>
         </start>
  <define name="Language">
    <element name="Language">
      <attribute name="Lan">
        <value>FrenchToDutch</value>
         </attribute>
        </element>
         </define>
        </grammar>
```

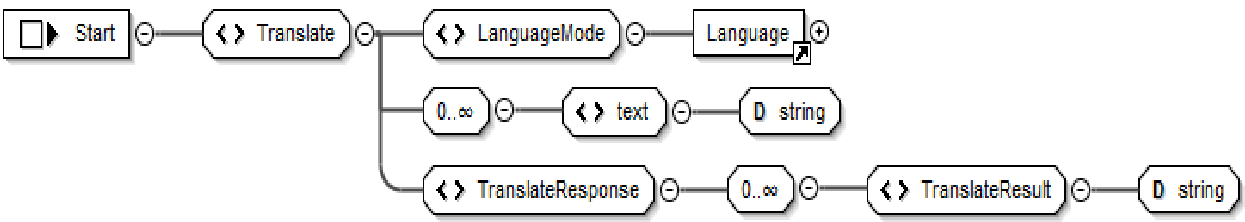

Fig 3: Directed Graph Representation of Schema Document translate.rng

C_1= Translate C_2 = LanguageMode, Text, TranslateResponse C_3 = TranslateResult

Fig 4: Listing of Equivalence Classes of translate.rng

(1) $$IIC = \sum_{i=1}^{3} \left(FOC_i^{\,2} / NE \right) + NA$$

$$= \left(\frac{1^2+3^2+1^2}{5}\right) + 1 = 3.2000$$

(ii) $$EF = \frac{e}{NE} = \frac{4}{5} = 0.8000$$

4. RESULT AND DISCUSSIONS

Series of experiments were conducted to show the effectiveness of schema language using IIC, as performance measurement. Analyses of all the implemented RNG can be seen in Table 1.

Table 1. Complexity Measures for RNG Schema Language

S/No	Schemas	DSERS	EF	IIC
1	Subsets	3.4000	0.9000	4.4000
2	Ping	3.7619	0.9523	3.7619
3	Saludar	2.7142	0.8571	2.7142
4	Translation	2.2000	0.8000	3.2000
5	ValidateCard	2.3333	0.8333	2.3333
6	Getbible	3.6315	0.9473	3.6315
7	Books	4.7088	0.9705	5.7088
8	AddressBook	1.0000	0.6666	3.0000
9	Authorization	3.3846	0.9230	4.3846
10	Mutants	2.8461	0.9230	2.8461
11	StockHeadlines	7.0000	0.9375	8.0000
12	ConvertTemp	3.0000	0.8333	3.0000
13	Links	5.0000	0.9090	7.0000
14	Phone	5.6666	0.8888	7.6666
15	World	4.1176	0.9411	4.1176
16	Advert	1.2692	0.9230	3.2692
17	GetData	2.7142	0.8571	5.7142
18	AccountExits	3.0000	0.9285	5.0000
19	PowerUnits	2.5000	0.8750	3.5000
20	Table	1.8000	0.8000	3.8000
21	Inventory	2.7142	0.8571	5.7142
22	GasMeter	3.3750	0.9375	3.3750
23	GetTariff	2.7142	0.8571	5.7142
24	Lot	3.2352	0.9411	3.2352
25	BonPlan	3.7692	0.9230	5.7692

26	LinearAds	3.1666	0.9166	4.1666
27	Variables	2.6666	0.9166	3.6666
28	Log	3.3750	0.9375	4.3750
29	Bank	1.6666	0.9375	3.6666
30	BlZServices	3.1818	0.6666	5.1818
31	Briefs	5.0000	0.9444	7.0000
32	CalServices	3.8000	0.9333	5.8000
33	Soap	1.5454	0.8000	2.5454
34	Contact	12.0232	0.9767	14.0232
35	ArendsogServices	11.8800	0.9600	13.8800
36	Account	11.8571	0.9714	11.8571
37	Collection	4.0000	0.9285	6.0000
38	VerifyRecord	2.8461	0.9193	2.8461
39	EmaiStmp	4.0000	0.9375	4.0000
40	FedACHcities	4.0000	0.9375	5.0000

4.1 Comparative Study of Existing Metric (DSERS), IIC Metric and EF Metric in RNG Schema Language

The graph in Figure 5 is inversely related because IIC had greater complexities values compared to DSERS i. e. the lower the values of DSERS, the higher the values of IIC. Thus, these have the same meaning of lower psychological complexity of RNGs. As a result more diversity of the elements (i.e. appearance of the elements in any order) increases the complexity values which showed more regularity and reusability traits and high frequency occurrence of similarly structured elements makes the developer more familiar with the schema language structure therefore; overall understandability of schema document becomes much easier because less understandability of schemas require more maintenance efforts.

The higher EF values for RNG in this graph could be interpreted as that elements were highly connected and dependent on each other. Thus, any modification made in any individual element could automatically update the other element to which that individual element is connected. Lastly, Figure 6 represents the virtual graph of comparison between IIC and DSERS metrics. The existing measure (DSERS) graph is the reverse of the IIC measure due to lack of many inheritance features; lack of number of attributes features of the schema documents, lack of similarly-structured elements and lack of higher frequencies of occurrences of the elements.

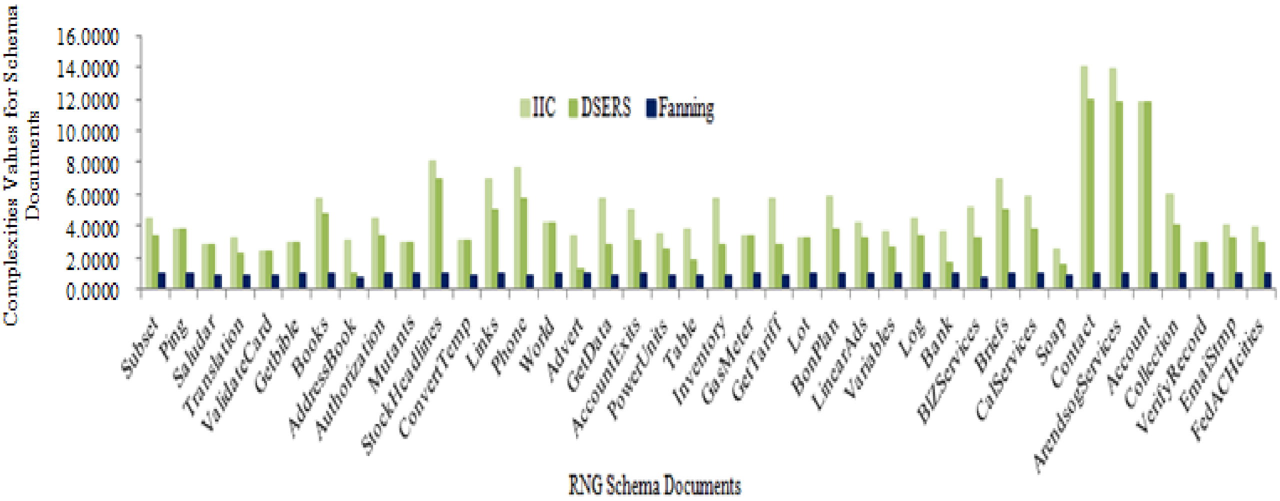
Complexities Values for Schema Documents
16.0000
14.0000
12.0000
10.0000
8.0000
6.0000
4.0000
2.0000
0.0000
IIC
DSERS
Fanning
Subset
Png
Sahidar
Translation
ValidateCard
Getbible
Books
AddressBook
Authorization
Mutants
StockHeadlines
ConvertTemp
Links
Phone
World
Advert
GetData
AccountExits
PowerUnits
Table
Inventory
GasMeter
GetTariff
Lot
BonPlan
LinearAds
Variables
Log
Bank
BIZServices
Briefs
CalServices
Soap
Contact
ArendsogServices
Account
Collection
VerifyRecord
EmaiSmp
FedACHcities
RNG Schema Documents

Fig 5: Graph of Existing Metric (DSERS), IIC Metric and EF Metric in RNG Schema Language

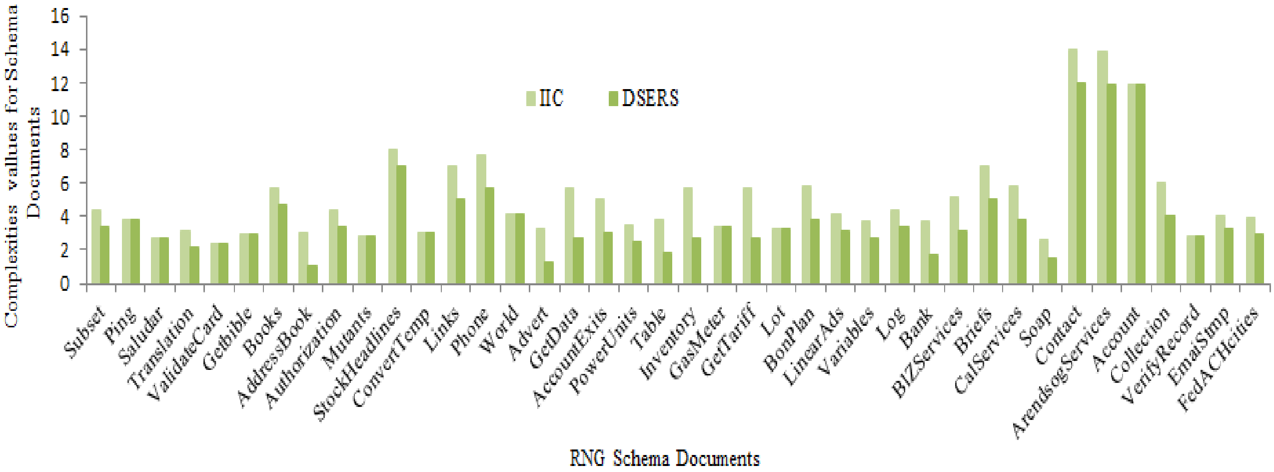

Fig.6: Comparison of Existing Metric (DSERS) and IIC Metric in RNG Schema Language

5. CONCLUSION

This paper work proposed Improved Interface Complexity (IIC) Metric based on information contained in the WSDL of the schema documents. The proposed metric considered fundamental factors which directly affect the complexity of the schema document, it used number of attributes (NA), number of equivalence class (NEC), frequency occurrence of the class (FOC), number of element (NE), and element fanning (EF), since information is contained in the elements and attributes of the schema documents. The study also presented in RNG schema language measure IIC techniques which were analyzed and the benefits features of the RNG schema language is discussed.

The IIC makes more sensitive measurement in understanding the information content contained in the schema documents. The applicability of the metric was evaluated by different schemas implemented in RNG to prove its robustness and effectiveness. The difficulties in understanding the schemas documents were measured and the results showed that RNG is a more suitable language when compared with existing measure as RNG was able to measure class elements comprehension, of a fact empirical validation have shown that RNG is able to reflect strong support for class elements to make them appear in any order. Lastly, RNG is highly structured and can partner with other schema language with a separate data typing language which makes it simpler in exhibiting a better presentation of a given schema document.

6. ACKNOWLEDGMENTS

Our thanks to the institution that provided us support and funding that contributed towards success of this work.

7. REFERENCES

[1] Jennifer N. R. (2006). Web Design in a Nutshell: A Desktop Quick Reference, *O'Rielly Amazon.Com ISBN*:9780596009878, Massachusetts College of Art Boston.

[2] 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.

[3] Thaw T. and Khin M. (2013). Measuring qualities of XML schema documents, *Journal of Software Engineering and Applications*, 6, 458-469.

[4] Sankar M. and Irudhyaraj A. (2014}. Software quality attributes for secured web applications. *International Journal of Engineering Science Invention*, 3(7):19-27.

[5] Romano D. and Pinzger M. (2011). Using Source Code Metrics to Predict Change-Prone Java Interfaces. *In Proceedings of ICSM'11,* (pp. 303–312).

[6] Dig D. and Johnson R. (2005). The Role of Refactoring in API Evolution. *IEEE International Conference on Software Maintenance, In Proceedings of ICSM, Budapest, Hungary,* (pp. 1–10).

[7] Mens T. and Tourwe T. (2004). A survey of software refactoring, *IEEE TSE,* 30(2), 126-138.

[8] Misra S. and Cafer F. (2011). Estimating complexity of programs in python language. *Technical Gazette,* 18(1), 23-32.

[9] Sotonwa K. A., Olabiyisi S. O. and Omidiora E. O (2014): Comparative analysis of software complexity for searching algorithm using code based metrics. *International Journal of Scientific & Engineering Research*, 4(6), 2983-2993.

[10] Makoto M., Dongwon L., and Murali M. (2001): Taxonomy of XML schema languages using formal language theory*, Extreme Markup Language:* XML source via XSL, Saxon and Apache FOP Mulberry Technologies, Inc., (pp. 153-166).

[11] Sotonwa K. A. Olusi T. and Olabiyisi S. O. (2019). Reviews of metrics on XML schema languages. *International Research Journal of Computer Science (RJCS),* 12(6), 700-708.

[12] Sotonwa K. (2020). Comparative analysis of XML schema languages for improved schema entropy. *FUOYE Journal of Engineering and Technology (FUOYEJET),* 5(1), 36-41.

[13] Marconi, M. and Nentwich C., eds. (2004). CLiX Language Specification Version 1.0.

[14] Clark J, Makoto, Murata, eds. (2001). RELAX NG Specification. OASIS. *See*http://www.oasis-open.org/committees/relax-ng/spec.html.

[15] Dongwon L. and Wesley W. C (2000). Comparative analysis of six XML schema languages, *In ACM SIGMOD* 29(3), 1-12.

[16] Makoto M., Dongwon L., Murali M. and Kohsuke K. (2005). Taxonomy of XML schema languages using formal language theory. *In ACM Transactions on Internet Technology (TOIT),* 5(4), 660-704.

[17] DeRose S. (1997). The SGML FAQ Book: Understanding the Foundation of HTML and XML. *Kluwer Academic Publishers. ACM Computing Surveys,* 31(4).

[18] Sotonwa K. A., Olabiyisi S. O., Omidiora E. O and Oyeleye C. A. (2019). SLOC metric in RNG schema documents. *International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)* 8(2), 1-5.

[19] Basci D. and Misra S. (2010). Entropy as a measure of quality of XML schema document, *International Arab Journal of Information Technology*, 8, 75-83.

[20] Basci D., Misra S. (2011). Document type definition (DTD) metrics, *Romanian Journal of Information Science and Technology,* 14(1), 31-50.

[21] 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 Center of ICT/ICE Research, CUCRID Building, (*pp. 1-10).

[22] Boxall M. and Araban S. (2004). Interface Metrics for Reusability Analysis of Components, *Proceedings of the 2004. Australian Software Engineering Conference,* Melbourne, Australia.

[23] Henry S. and Kafura D. (1981) Software structure metrics based on information flow, *Computer Journal of IEEE Transactions on Software Engineering*, 7(5), 510-518.