FORMALIZATION OF VERSIONING RULES FOR XML SCHEMA USING UML CLASS DIAGRAM

1HANANNI AMAN, 2ROSZIATI IBRAHIM

1Department of Software Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), MALAYSIA
2Department of Software Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), MALAYSIA

E-mail: 1hanani@uthm.edu.my, 2rosziati@uthm.edu.my

ABSTRACT

In agile software development methodology, XML schema is used for developing the web applications. The major problem in using the agile software development methodology is capturing the software requirements especially when versioning occurred in XML Schema. This paper presents how to capture software requirements for document changes when versioning occurred in XML Schema. UML Class Diagram is used in addressing the versioning rules. Once the versioning rules are captured, these rules are then being formalized for better understanding of the versioning problem occurred in XML Schema.

Keywords: XML Schema, Versioning Rules, Traceability, UML Class Diagram, Document Changes

1. INTRODUCTION

XML is very popular in web community in sharing and transferring data between applications. The document and schema are multiplying and it needs to be maintained. Management of XML schema and document is very tedious and time consuming because of the XML nature itself. XML commercial tools offer editing XML document and schemas, specify validation in the document, organize and manage related files. Some tools assist with nice graphical interface, highlight validation error and view xml in tree structure. If any editing or updating of the file, every tool will keep the final editing as a file. However, any changes in XML schema or document cannot be traced after the file is saved.

Keeping changes is one of the important criteria in managing a file. The idea of keeping the changes in version concept arises when XML needs to keep track of the changes. This is because any changes occurred may change any requirements for the designer and end user. Hence the software developed based on it will have a different version. Software versioning is a process used in agile development to give unique names or numbers when changes occur in state of the software. The number assigned to the version is in increasing order within a given version number category based on major or minor new developments. Underlyng of versioning process is how to handle any changes occurred. Hence this paper addresses the issues of versioning rules. The rules for versioning are captured first. Class diagram is used to capture the changes occurred in XML Schema. Then the versioning rules are being formalized. The formal method is used for the formalization of versioning rules. The formal rules are then being tested using the case study to see the effectiveness of the versioning rules for the traceability link between the two class diagrams.

The rest of the paper is organized as follows: Section 2 is the related work regarding the research areas and Section 3 is the preliminaries discussion regarding the XML Schema and UML Class Diagram. Section 4 presents the versioning rules for the XML Schema and Section 5 discusses the formalization of the versioning rules. The case study is presented in Section 6 and finally, Section 7 is the conclusion of this paper.

2. RELATED WORK

Keeping changes in XML starts when Tan & Goh [1] suggest a need to highlight changes and differences between a preceding version or a variant and the original standard of XML. The authors also emphasize when new functionalities establish, preceding XML need to be kept for maintenance purpose. It is about compatibility issue with XML based standard. As the XML standard evolves, Cavaleri et al. [2] focus on updated
schema effect when a XML standard, XQuery is applied using a language called eXupdate [3]. Finally, keeping primitive changes of XML files is proposed by Brahmia et al. [4]. Their paper focused on Part 0 in W3C recommendation group. However, every change is based on the XML schema only. Any changes effect on conceptual level is not shown in previous research. Other researchers [5, 6] have seen the bigger picture of evolving and managing XML schema using class diagram change requirements. The change is on the conceptual model and makes the XML schema evolves and changes. However, forward direction does not show the reality of XML changes impact on conceptual models.

In software engineering area, more researchers focus on forward engineering of XML which is generating XML document from conceptual model [7, 8, 9, 10, 11, 12 and 13]. Forward engineering is a method to generate XML schema from conceptual model. However, research on generating conceptual model from XML application gain less focus. The reverse engineering for XML is also important as forward engineering because it is an alternative way to assist Agile developer generates conceptual models for communication and documentation purposes.

In software development process, XML schema modification needs to be maintained consistently by using the documentation method. The XML schema modification involves textual changes caused by the changes issue such as schema style changes [14], requirement changes [12, 7, 14, 15], adaptation to new XML technology [2], and integration with other XML schema [4]. The changes impact needs to be understood by end users, designers and software developers. Therefore, this paper addresses the issue of changes in XML documents by formalizing some of the important rules for document versioning (that is when changes occurred). To see the effectiveness of the versioning rules, a case study is also presented in this paper for evaluation of the versioning rules.

3. PRELIMINARIES

This section describes two important factors in formalizing the versioning rules: XML Schema and UML Class Diagram. The XML Schema is used to address the versioning rules in web applications. UML Class Diagram is used for traceability link to detect document changes for XML Schema. XML Schema and Class Diagram are used to formalize the versioning rules. Versioning occurs when there are changes in a XML schema based from the previous XML schema.

2.1 XML Schema

XML Schema is used for developing web applications. It has rigorous specification of XML document. XML schema consists of rules that define the constraint of each element in the XML document which provide a mean for defining the structure, content and semantics of XML documents in more detail [16]. If any element and attributes in XML document satisfy the constraint in the schema, a valid report will be generated.

XML schema consists of 5 components [16, 17]. They are type definition, element declaration, attribute declaration, attribute group definition and model group definition. Table 1 gives the definitions of XML schema components.

<table>
<thead>
<tr>
<th>XML Schema components</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type definitions</td>
<td>To constraint the structure of element and attribute used in XML document. Consist of simple type definition and complex type definition.</td>
</tr>
<tr>
<td>Element declaration</td>
<td>An association of a name with a type definition, either simple or complex. It also has an attributes within. Each element may occur more than once in the schema.</td>
</tr>
<tr>
<td>Attribute declaration</td>
<td>An association between a name and a simple type definition, together with occurrence information.</td>
</tr>
<tr>
<td>Attribute group definitions</td>
<td>A group of attribute declaration with a name of the group. This group is embedded in element together in complex type definition or element or with attribute in an element.</td>
</tr>
<tr>
<td>Model group definition</td>
<td>An association between a name and a model group, enabling re-use of the same set in several complex type definitions</td>
</tr>
</tbody>
</table>

An XML schema component is a primary component in setting element in XML document. In addition of that, elements need to be relating to other XML schema components. Figure 1 shows the example of XML schema.
Based from Figure 1, the 5 components in XML Schema are used using the tagging. The XML Schema shows that one element named <book> contains attribute named <isbn> with string type. Other element named <title>, <author>, and <character> is nested in the <book> element. Each <title> and <author> element has string type. While <character> element has a set of element named <name>, <friend-of>, <since> and <qualification> nested with it. <name>, <friend-of> and <qualification> element also use string type. Meanwhile, <since> element uses date type. <character> element may occur within <book> element either no element or maximum unbounded. Occurrences for other element that nested to <book> element, which is <title> and <author> occurs once only. <friend-of> element occur either no element or maximum unbounded relate to <character> element. Based on this example, XML Schema can be understood easily. However using this format of the XML Schema presentation is difficult for the user to review the contents of the application.

2.2 UML Class Diagram

UML (Unified Modelling Language) [18] Class Diagram is used during design phase of software development life cycle (SDLC). Under SDLC four main phases are used when developing any software or system. They are analysis, design, implementation and testing. During design phase, software specification will be developed using UML. UML specified 12 diagrams which include use-case diagram, class diagram and others. The class diagram is the main static analysis diagram [18]. It shows the static structure of the model for the classes and their relationships. They are connected to each other as a graph. Each class has its own internal structures and its relationships with other classes. Figure 2 shows an example of a class diagram for Monitoring System of Postgraduate Students.

Based from Figure 2, classes FocusGroup and PostGraduateStudent are subclasses from class Person. Class PgForm has 4 functions (GetForm, Submit, View and Evaluate). Further details regarding subclasses and mapping of classes and subclasses functions with use cases can be found in [19].
4. THE VERSIONING RULES (VR)

As stated earlier in this paper, changes occurred in XML documents need to be understood and documented. The issue of changes in XML documents is addressed by using the versioning of the XML documents. If XML document changes, it implies that the versioning occurs. Therefore, to address the versioning issue, two class diagrams are used from XML schemas for versioning purposes.

The earlier XML Schema is used to generate a class diagram called preceding class diagram. When maintenance process occurs, new XML schema is introduced. This new XML schema has been modified because of the number of XML schema components have been changes. The changes are visible in syntax. In order to capture the semantic changes of the new schema, the new schema need to be transformed into class diagram. The same transformation is used to generate the preceding class diagram. The modified preceding class diagram is now called a version class diagram. The version class diagram needs to be compared with preceding class diagram to capture the additional, removal and rename of class diagram components. Figure 3 shows the scenario of this.

Figure 3: Traceability Link Between Two Class Diagrams

Based from Figure 3, a single change of XML schema impacts the specification of the requirement document too. The manual update of requirement changes is tedious and laborious. Therefore, when XML schema’s new version is introduced, the requirement document needs to be updated. When requirement changed, number of classes in version class also changed. In this research, a class is changed when number of version classes is more than classes in preceding class diagram. Figure 4 shows the versioning occurred between two XML Schemas. Therefore, in order to detect the changes between two class diagrams, few steps are performed for traceability factor. Table 2 shows the three steps performed based from traceability link in Figure 4.

<table>
<thead>
<tr>
<th>Changes</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Each class of both class diagrams has been compared - level by level of the diagrams.</td>
</tr>
<tr>
<td>Step 2</td>
<td>Each Added classes and Removed classes are compared.</td>
</tr>
<tr>
<td>Step 3</td>
<td>Classes which have the same attribute and same relationship are tested with string similarity. High result of similarity tests the changes of new versions of class diagram.</td>
</tr>
</tbody>
</table>

Based from Table 2, the modification is generated by three (3) atomic operations classified which are additional, removal and rename operations of XML basic components (element, attribute, type definition, model group and attribute group). An additional change involves in adding any element or attribute in the schema whereas a removal change involves any removing element of XML components. Beside of that, rename operation occurs when an element’s name is given new name and maintain the set of attributes and relationship. A rename operation involves only one (1) class that has been renamed. Combination operation involves combinations of these three (3) atomic operations. These combinations will generate new semantic changes such as migration of element to another element where involves removal and additional at different level of class diagram. But in this research, a new kind of semantic changes is proposed. The proposed combination changes occur when a removed element from preceding class diagram is added by two (2) new elements which maintain the same set of attribute and relationship as the removal element. This link is called as traceability link.

In order to show the new combination of atomic operations, the Versioning Rules (VR) is introduced. It involves additional, removal, rename and combination of these operations on element. Tracing changes operations occur on class diagram as class diagram represents semantic changes of XML schema versioning.

**Versioning Rule 1:**

There are two (2) class diagrams that need to be transformed from XML schemas for versioning. The new version is a subset of preceding schemas. The newer versions has maintained some of classes, attributes, relationship from the preceding class diagram, moreover new classes are added and some classes are removed.
5. FORMALIZATION OF VERSIONING RULES (VR)

In order to formalize the versioning rules, we need formal definitions of class diagram and transformation rules. These formal definitions are already presented in our previous paper in [20]. For consistency, we presented part of it here in order to establish the versioning rules formula.

Rule 1: Class diagram consists of:
- Class
- Relationship

where
- Class is consists of a set of attributes and each attribute may have multiplicity.
- Relationship is a connection between classes. Each association may have multiplicity. There are three (3) kinds of relationships which are association, aggregation and dependency.
  - Association is a connection between classes that relates.
  - Aggregation is a connection between classes where a class has to be a part of other class.
  - Dependency is a connection between classes where a class are dependent on other class.

The formal definition of class diagram is as follow:

Definition 1: Let “cd” be a class diagram. “cd” consists of a finite set of classes and a finite set of relationships. Thus, class diagram can be defined as

\[ cd = \{CLS, ASSOC\} \]

where

\[ CLS = \{cls_i | 1 \leq i \leq n\} \text{ is a finite set of classes} \]
\[ ASSOC = \{assoc_i | 1 \leq i \leq n\} \text{ is a finite set of relationships} \]

The class has a set of attribute and attribute’s type. Each attribute has a multiplicity value to show the number of occurrences appear in each class. The class definition is illustrated in Definition 2.

Definition 2: Let “cls \in CLS” be classes consist in a class diagram. The class consists of a finite set of attribute and relationship of a class. Thus class can be defined as

\[ cls_i = \{\text{attr}_1, T_1, \ldots, \text{attr}_n, T_n\} \]

where

\[ \text{attr}(cls_i) \text{ is a set of attributes of class and} \]
\[ T_i \text{ denoted to type of attribute, attr}_i \text{ of a class, cls}_i \text{ with a multiplicity value} \]
\[ \text{where “multiplicity” be a multiplicity of classes or attribute.} \]

Multiplicity consists of value to show the number of minimum and maximum occurrences of association class or attribute in each class. Thus multiplicity can be defined as

\[ \text{multiplicity (cls)} \subseteq [1, \infty] \]

Classes relationship is defined as “assoc” where each “assoc” may has one of these values, which are association, aggregation and dependency as stated in Definition 3.

Definition 3: Let “assoc” be an association between classes. Thus it can be defined as unary relationship of class.

\[ CLS \times \ldots \times CLS \]

where

\[ assoc = \{cls_i, cls_j\} \]

where

\[ t(assoc) = \{\text{association, aggregation, dependency}\} \]
**Definition 4.** Let “cdv” be a version of preceding class diagram, “cd”. “cdv” has more number of classes than classes of preceding class diagram, “cd”. Thus, cdv can be defined as

\[ |cls(cdv)| > |cls(cd)| \]  

(4)

Based on Definition 4, versioning rules are defined. Versioning rules define a traceability factor by comparing two class diagrams. Changes will be recognized as added when there are new classes in version class diagram and removed class when there are missing classes in version class diagram. The versioning rules are shown in VR1, VR2 and VR3.

**VR1.** Let “cdv” be an added factor of class diagrams where a new class, “\(\text{cls}_j\)”, is found in version class diagram, “cdv” as compare with class diagram, “cd”. Thus

\[ cd \setminus cdv = \text{cls}_j \]  

(5)

**VR2.** Let “cdv” be a removed factor of class diagrams where class, “\(\text{cls}_i\)”, is missing in version class diagram, “cdv” as compare with class diagram, “cd”. Thus

\[ cdv \setminus cd = \text{cls}_i \]  

(6)

The removed class will be the target of next changes detection. VR1 is used for the rule where the change detection steps are listing the added classes on new version of class diagram and VR2 is used when the listing are removed from the removed classes from preceding class diagram, where two (2) added classes must relate with one (1) removed class. The related components between class diagrams are classes and versioning criteria are assoc and attribute which are presented in VR3.

**VR3.** Let “\(\Delta(\text{cls}_i)\)” be a traceability factor of class diagrams where a classes, “\(\text{cls}_i\)”, “\(\text{cls}_j\)”, and “\(\text{cls}_k\)” of version class diagram, “cdv” is evolved from a class, “\(\text{cls}_i\)”, of preceding class diagram, “cd” if “\(\text{cls}_j\)” and “\(\text{cls}_k\)” of version class diagram, “cdv” has the same “\(\text{assoc}(\text{cls}_i)\)” and “\(\text{assoc}(\text{cls}_j)\)” of version class diagram, “cdv” is the same with “\(\text{assoc}(\text{cls}_i)\)” from preceding class diagram, “cd”. Thus,

\[ \Delta(\text{cls}_i) = \begin{cases} 1 & \text{if } \alpha_i = \beta_i \land \text{assoc}(\text{cls}_i) = 0 \\ 0 & \text{otherwise} \end{cases} \]  

(7)

where

\[ \alpha_i = \beta_i \land \text{assoc}(\text{cls}_i) \]

Based from the versioning rules (VR1, VR2 and VR3), any changes detected in version class diagram from preceding class diagram will be marked and traceability factor will become 1 (one). If there are no changes, then the traceability will be 0 (zero).

Under SDLC, the next step after formal specification is implementation phase. We implement the tool based on the formal rules for these versioning rules and transformation rules [20]. The rules are embedded within the tool to make the detection of changes easier. The details of the tool named XML-DocTracker can be found in [21]. The purpose of implementing the SML-DocTracker is to being able to generate Software Requirements Specification (SRS) automatically from the XML Schema.

**6. A CASE STUDY**

In this section a case study is presented to show how the versioning rules are being used for the traceability link and to generate software requirements for document changes. The transformation rules are used from [22].

Case study has been taken from Liquid Technology web entitled Electronic Business using eXtensible Markup Language (ebXML) [23]. This case study based on ebXm schema involved of all the XML construct definition (element, attribute, attributeGroup, complexType, model group and simpleType) titled cppinga.xsd for version 1 and ccpp-
cpa-2_0.xsd for version 2 [23]. ebXML is an open specification of XML based infrastructure which provide global use of electronic business information for an interoperable, secure and consistent manner sponsored by Organization for the Advancement of Structured Information Standards (OASIS) and UN/CEFACT. UN/CEFACT is an organization that makes International Electronic Data Interchange (EDI) standards for electronic trade documents in XML format [23].

Table 3 shows the number of components used between two (2) versions of ebXML schema. This design allows reuse of both elements and types. The design benefits the high reuse of class. However, this kind of schema is difficult to read and understand. The difficulties make this case study suitable to address the versioning problems.

Table 3: Ebxml Schema Component Analysis

<table>
<thead>
<tr>
<th>XML schema component</th>
<th>Cppa schema</th>
<th>Cpp-cpa-2.0 schema</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Element</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td>Global Attribute</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Local Element</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Global Complex Type</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Global Simple Type</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Group Model</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Attribute Group</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Element with Ref</td>
<td>43</td>
<td>53</td>
</tr>
</tbody>
</table>

From Table 3, the XML schemas are used to generate the class diagrams. Table 4 shows the analysis for the two class diagrams from the earlier XML Schema version (Cppa) and the later version (Cpp-cpa-2.0).

Table 4: Ebxml Class Diagram Analysis

<table>
<thead>
<tr>
<th>XML Schema</th>
<th>Cppa</th>
<th>Cpp-cpa-2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>48</td>
<td>62</td>
</tr>
<tr>
<td>Association</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>Aggregation</td>
<td>43</td>
<td>75</td>
</tr>
<tr>
<td>Dependency</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4 shows the number of components used between two (2) versions of ebXML schema. This design allows reuse of both elements and types. The design benefits the high reuse of class.

We then used this case study to evaluate the XML-DocTracker [21] for its effectiveness in tracking the changes. The evaluation of XML-DocTracker is compared with Necasky in [11]. XML Schema components of both schemas are transformed based on reverse transformation rules [22] using XML_DocTracker tool [21]. As for evaluation, this reverse method is evaluated and compared with [11]. The evaluation results are shown in Table 5.

Table 5: Comparison Of Evaluation Of Two Methods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Class</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td>Association</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Aggregation</td>
<td>43</td>
<td>n/a</td>
</tr>
<tr>
<td>Dependency</td>
<td>0</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Based from Table 5, it is shown that XML-DocTracker is able to transform more classes with aggregation relationship. This more specific relationship is then being used for generating the new Software Requirements Specification [21]. Note that in earlier research done by Necasky [11], the limitations occurred for aggregation and dependency. They do not address aggregation and dependency issues. Our research added these two issues in tracking the changes in XML Schema.

7. CONCLUSION AND FUTURE WORK

In this paper, we have presented a formalization of versioning rules for XML Schema using class diagram. The versioning rules are used for traceability factors in order to detect document changes in XML Schema. The changes of XML documents are not easy to keep track manually. Therefore, giving the versioning rules for the XML document changes help to solve the document changes issue. However, since the informal rules of versioning rules do not have formal definition, we formalized the rules for better understanding.

The tool named XML-DocTracker has been implemented based on the transformation rules and versioning rules in order to regenerate the documents (Software Requirements Specification) from the XML Schema.

For future research, the theorem prover can be used to test the correctness of the formalized rules. Operational semantics can also be used for the proof theory of the formalized rules.
ACKNOWLEDGEMENTS
The authors would like to thanks Malaysian Ministry of Higher Education (MoHE) for supporting this research under the Fundamental Research Grant Schema (FRGS).

REFERENCES:


Figure 4: Traceability Link between two class diagrams from two XML Schema when Versioning Occurred