I. J. Computer Network and Information Security, 2019, 12, 30-39
Published Online December 2019 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2019.12.04

Evaluation of Quality for Semi-Structured Database System

Rita Ganguly
[1] Department of Computer Applications; Dr.B.C.Roy Engineering College; Durgapur: 713206; West Bengal; India
E-mail: ganguly.rita@gmail.com

Anirban Sarkar
[2] Department of Computer Science; National Institute of Technology; Durgapur;713209; West Bengal; India
E-mail: sarkar.anirban@gmail.com

Received: 25 October 2019; Accepted: 17 November 2019; Published: 08 December 2019

***Abstract*—**The quality evaluation of transactional calculus for semi-structured database system develops metrics for data quality. A conceptual data model of higher quality leads to a higher quality information system. Quality of a data model may affect with effectiveness i.e. the quality of results and the efficiency like time, effort etc. of development of information system. Hence, boosting up the data model quality is also prone to improve quality of delivered system. An array of metrics for quality has been proposed for the semi-structured data model with proper blend of metrics framework suitable for transactional calculus for data model of semi-structured. This paper proposes a framework for quality evaluation of transactional calculus for semi-structured database system using TCSS X-Query. In the proposed quality evaluation, the viewpoint has been described using a set of proposed quality measurements. Each of these quality measurements is linked with set of related metrics. The framework comprised of direct and indirect metrics for the purpose of quality evaluation. The framework facilities a double-fold view point using a set of quality measurement. In quality evaluation two viewpoint quality dimensions are focused: like designer level viewpoint and user level viewpoint. The proposed metrics set and measurements have been validated empirically. The purpose of empirical validation is to establish the metrics are practically useful for the assessment of quality measurements and operability factor.

***Index Terms*—**TCSS, Semi-structured, Metrics, Empirical Validation, GQL-SS, Quality Evaluation.

I. Introduction

Leading a crucial role of semi-structured database system using by large number of data processing application on web. In recent days the semi-structured data has became prevalent with the growing demand of such internet based software system. The formal structure of data is not strictly conforming by semi-structured data. Rather it posses irregular and partial organization. Semi-structured data is data that is neither raw data, nor very strictly type as in conventional database systems. Semi-structured data is also known as self-describing structure as it contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. It doesn't conform other forms of data tables or formal structured of data models associated with relational database.

Only on a good conceptual model the designing of efficient database of semi-structured may be done. A good conceptual level data model of semi-structured database must accompany with semantic enriched detailed operational model for querying semi-structured data. XML technology serves as de facto standard for representing at the logical level semi-structured data and X-Query techniques for querying such data. But still there is no widely acceptable formal methodologies exist for those techniques.

An algebra based query language called GQL-SS (Graphical Query Language for Semi-Structured) has been proposed for management of semi-structured database. The proposal also supports efficient navigation over such data represented as the logical level XML technology. The proposed query algebra has been defined based on the Graph Object Oriented Semi-Structured Data Model (GOOSSDM) described in [2].

Traditional database system forces all data to adhere to an explicitly specified, rigid schema and most of the limitations of traditional database may be overcome by semi-structured database. Neither semi-structured data is very strict type like as conventional database systems nor a raw data. Whereas a traditional transaction system guarantee that either all modifications are done or none of these i.e. the database must be atomic in nature. To overcome traditional transactional problems, extending the transactional processing system in semi-structured database with consistency, availability, and partition tolerance (CAP)[12] coordination, with basic availability, soft state and eventually consistent (BASE)theorem with new healthiness conditions, enriched with a standard designing model. The motivation of the Transactional Calculus for Semi-structured System proposes a transactions framework based on mathematical expression

(transactional calculus for graph semantic based semi-structured data model (TCSS)) where a transaction is treating as a compensable program mapping from its environment and provides a transactional calculus of refinement. It proposes to show that most of the semi-structured transaction can be converted to a calculus based model which is simply consisting of activities like forward and a compensation module of CAP and BASE theorem. It is important that the service still perform as expected when some nodes crash or communication links fail. Besides, the paper includes verification of several useful properties of the proposed TCSS. Moreover, a detailed comparative analysis has been providing towards evaluation of the proposed TCSS.

A quality evaluation framework has been described with a viewpoint for the semi-structured data model. A higher quality information system will be lead by a higher quality conceptual data model. The performance (quality of results) and productiveness (time, cost, effort) of information system development may be affected by the quality of conceptual data model. Hence, enhancing the level of conceptual data models is also likely to improve quality of delivered systems, which is also true for the transactional calculus for semi-structured database system.

The purpose of the quality evaluation are as follows-(1) for whom and why it is undertaken (2) at a particular point in time why the evaluation is being undertaken , (3) for acquisition of skills and/or liability functions, how the evaluation is to be used. For example the evaluation's overall purpose may be to: (1) continuation or discontinuation of a project or program (2) introducing a program /development procedure (3) skills for developing of specified results. Specific goal of the evaluation is: what the evaluation aims to find out that clarified by the specific objectives of the evaluation. For example to: (1) in order to bring attention for future design and implementation with respect to a specific development intervention provide finding, conclusion and recommendations.(2) assessing the effectiveness, efficiency , relevance and sustainability of a specific development intervention and ascertain results like output, outcome ,impact etc. The scope of the quality evaluation is: the development intervention being evaluated (the evaluation object) is clearly defined, including a description of the intervention logic or theory. Discrepancies between the planned and actual implementation of the development intervention are identified. Here, in the proposed quality evaluation, the viewpoint has been described using a set a proposed quality measurements. Related metrics are associated with each of the quality measurements. In quality evaluation, two viewpoint quality dimensions are focused like designer level viewpoint and user level viewpoint. The designer level viewpoint metrics are transaction/query throughput and query performance and user level viewpoint metrics are effectiveness and analyzability. The empirical validation process shows that, the several proposed metrics are practically useful for the estimation of feasible factor of conceptual level data and further there exist significant correlations among the feasible factor and measurements of proposed quality evaluation. Section II introduces the related previous work. Section III describes about the GOOSSDM. Section IV introduces the proposed set of direct and indirect metrics. Section V represents the implementation of proposed TCSS with implementation of TCSS X-Query and experimental results. Section VI introduces quality evaluation framework. Section VII presents empirical validation of proposed metrics with experimental settings and experimental steps. At last section VIII concluded.

II. Related Work

In previous work [27], it has been shown that most of the transaction of semi-structured data can be converted to a calculus based model which consists of a forward activity and a compensation module of CAP [12] and BASE [25] theorem. More precisely (i) described GOOSSDM [2, 19, 20, 21] schema and GQL-SS [11] data are associated to leaves as abstraction of the content of leaves may contain data variables of the path expression. They may also contain evaluation of the empty path or length of n edges path and preserving of labels or tags by using path expression. (ii) Developing of three types of algorithms: Three types of algorithms use to evaluate the path in GOOSSDM schema, one for searching return node, second for searching the path from root of GOOSSDM schema to the desired node and the third one is for the searching and listing of the tail nodes. (iii) Define the GQL-SS algebra for GOOSSDM model that operate on semi-structured schema concept and / or several constructs described in the model. The algebra consists of a set of operators and few of them can be using with the constructs like ESG, CSG separately. Lastly (iv) Define a transactional calculus for GQL-SS Algebra consists of a set of operators and proving the queries by using X-Query in TCSS. In this case, since dealing with the combination of CAP and BASE theorem, this proposal for expressing and executing queries and real time applications shown by using the calculus. Defining a topography language to plot attributes of the data sources to the global schema and bridge query language to write the calculus.

III. Illustration of GOOSSDM

Extending the object-oriented paradigm to semi-structured data model, the GOOSDM introduced. It's specifying the irregular and heterogeneous structure, hierarchical and non-hierarchical relations, n – array relationships, cardinality and participation constraint of instances with all details that are required for semi-structured data model. The entire semi-structured database to be viewing as a Graph (V, E) in layered organization that is allowed by the proposed data model (GOOSSDM).At the lowest layer, each vertex represents an occurrence of an attribute or a data item.

Let consider an example of Project Management System (PMS),[11], associated with Project. Project has attributes like members, department and publications. Several members are associated with project and each

member can participated in any project. Department contains member, and each individual members may have or not have publication. The PMS is semi-structured is in nature. The GOOSSDM schema for PMS has been shown in figure 1. The sample data (data are taken anonymously) is showing in Table I.

Table 1 Sample data set for PMS

Project 1									
Pname	**PID**	**Topics**	**Member**			**Department**		**Publication**	
			MID	**MName**	**Maddress**	**DID**	**DeptName**	**PuID**	**Ptopics**
ABC	P1001	AAAA	M01	Bipin	XX	D01	CSE	P001	RRR
XYZ	P1003	CCCC	M03	Ashu	PP	D02	CA	P003	SSS
DEF	P1004	DDDD	M04	Rashi	YY	D03	EE	P004	TTT
XYZ	P1005	QQQQ	M06	Sashi	RR	D03	EE	P005	VVV
ABC	P1001	BBBB	M07	Priya	CC	D01	CSE	P006	MMM
Project 2									
Pname	**PID**	**Topics**	**Member**			**Department**		**Publication**	
			MID	**MName**	**Maddress**	**DID**	**DeptName**	**PuID**	**Ptopics**
PQR	P1006	YYYY	M07	Priya	CC	D02	CA	P007	NNN

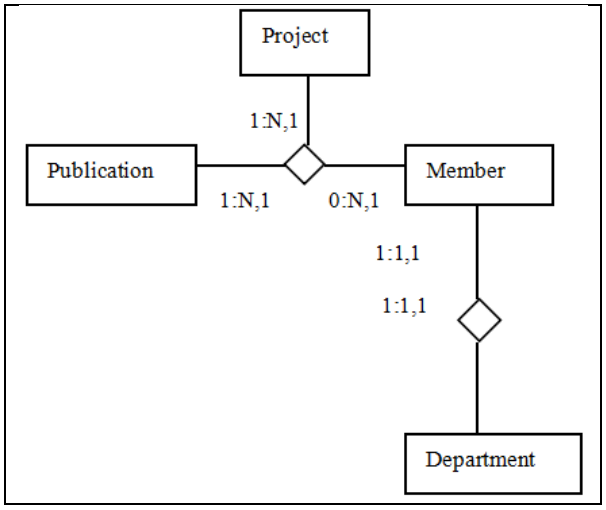

Fig.1. GOOSSDM Schema for PMS

IV. Set of Metrics for Propose Transactional Calculus of Semi-Structured

In this section, a group of construct level metrics has been proposed on schema of viewpoint of a semi-structured data model. Helping the measurement process, the mapping of the model should also be supplemented with a model of the mapping domain. A model specifies how the characteristics relate and how the entities are related to the attributes.

For pertaining to performance there are two important categories of metrics, direct metrics which are measured without the involvement of other entity and indirect metrics which are measured with the involvement of others. Further a separate set of construct level metrics has been proposed on unit, basis and analyzing point. Finally, for validation a set of metrics has been proposed for the data model of semi-structured database system.

A. Direct Metrics of Transactional Calculus for Semi-Structured Database System

The multidimensional schema is a representation of the concepts of TCSS application domain using a well-defined set of data model constructs. Direct metrics determine the top level health of database by measuring its useful output. Considering S is a schema for TCSS, the proposed direct metrics may be categorized into four subtypes, are:

- Throughput [Th(S)]: Throughput represents the amount of work which is done by the system per unit of time. Also indicates an application can handle how many no of transactions per second.
- Success[S(S)]: Success rate is calculated by dividing the total no of successful (approved) transactions by the total no of attempted transactions over a given time period. For examples, if you ran 100 transactions or queries, and 93 of them were successful, then transaction success rate is 93.The number of successful queries means the successfully executed works percentage.
- Error [E(S)]: Error rate is calculated by dividing the total no of unsuccessful (unapproved) transactions by the total no of attempted transactions over a given time period. The rate of errors per unit of time usually expressed as the number of erroneous results.
- Performance [P(S)]: How efficiently a component is doing its work that quantifies performance. Latency is the most common performance metric, which represents the time required to complete a unit of work. “99% of request returned within 0.1S”, that is expressed as an average or percentile of the latency.

B. Indirect Metrics of Transactional Calculus for Semi-Structured Database System

Indirect metrics are the required components of the database for successfully completion of the job or specified task. Some low level resources of indirect metrics are tangible components like memory, CPU, disks and n/w interfaces .Higher level resources such as query cache and database waits can be considered a resource and therefore monitored. Considering S is a schema for TCSS,

the set of proposed indirect metrics are-

- Utilization [U(S)]: How much database is involved in that is the percentage of time, or the using database volume in percentage.
- Saturation [Sa(S)]: Determination of the amount of the requested work that the database can't yet service, and waits in the queue.
- Availability [A(S)]: The database responded to request which is denoted as the percentage of the time.

V. An Implementation of Proposed TCSS

To introduce the syntaxes and semantics of TCSS, let us starting with a simple example of transactional query by using x-query. In this section, the expressiveness capabilities of the proposed transactional calculus of TCSS demonstrated by applying the calculus to suitable example queries. As for example the demo1.xml (PMS data in xml shown in Table 1) is given below:

```
<project>
<project1>
<pname>ABC</pname>
<pid>P1001</pid>
<topics>AAAA</topics>
<member>
<mid>M01</mid>
<mname>BIPIN</mname>
<maddress>xx</maddress>
<department>
<did>D01</did>
<dname>CSE</dname>
<publication>
<puid>P001</puid>
<ptopics>RRR</ptopics>
</publication>
</department>
</member>
```

```
<pname>XYZ</pname>
<pid>P1003</pid>
<topics>CCCC</topics>
<member>
<mid>M03</mid>
<mname>ASHU</mname>
<maddress>PP</maddress>
<department>
<did>D02</did>
<dname>CA</dname>
<publication>
<puid>P003</puid>
<ptopics>SSS</ptopics>
</publication>
</department>
</member>
<pname>DEF</pname>
<pid>P1004</pid>
<topics>DDDD</topics>
<member>
<mid>M04</mid>
<mname>RASHI</mname>
<maddress>YY</maddress>
<department>
<did>D03</did>
<dname>EE</dname>
<publication>
<puid>P004</puid>
<ptopics>TTT</ptopics>
</publication>
</department>
</member>
<pname>XYZ</pname>
<pid>P1005</pid>
<topics>QQQQ</topics>
<member>
<mid>M06</mid>
<mname>SASHI</mname>
<maddress>RR</maddress>
<department>
<did>D03</did>
<dname>EE</dname>
<publication>
<puid>P005</puid>
<ptopics>VVV</ptopics>
</publication>
</department>
</member>
<pname>ABC</pname>
<pid>P1001</pid>
<topics>BBBB</topics>
<member>
<mid>M07</mid>
<mname>PRIYA</mname>
<mid>M07</mid>
<mname>PRIYA</mname>
<maddress>CC</maddress>
<department>
<did>D01</did>
<dname>CSE</dname>
<publication>
<puid>P006</puid>
<ptopics>MMM</ptopics>
</publication>
</department>
</member>
</project1>
<project2>
<pname>PQR</pname>
<pid>P1006</pid>
<topics>YYYY</topics>
<member>
<mid>M07</mid>
<mname>PRIYA</mname>
<maddress>cc</maddress>
<department>
<did>D02</did>
<dname>CA</dname>
-<publication>
<puid>P007</puid>
<ptopics>NNN</ptopics>
</publication>
</department>
</member>
</project2>
</project>
```

1. ***Find the project name and project id from the CSG project1.***

```
for $p1 in doc("demo1.xml")//project1
for $p2 in doc("demo1.xml")//project1
where $p1//topics != $p2//topics
```

return<table ID="project">
<pname>{data($p1//pname)}</pname>
<pid>{data($p1//pid)}</pid>
</project>
</table>
<table ID=" project">
<pname> ABC XYZ DEF XYZ ABC
</pname>
<pid> P1001 P1003 P1004 P1005 P1001
</pid>
</project>
</table>

2. ***Find the details of publication whose Member Id MID="M03" and Publication Id PID="P003".***

for $p in doc("demo1.xml")//member
where $p//mid = "M03"
and $p//puid = "P003"
return $p//publication
<publication>
<puid> P003 </puid>
<ptopics> SSS</ptopics>
</publication>

3. ***Find the details of member where MName="Bipin" from project1 and also find the details of Member where MName="Priya" from Project2.***

for $p1 in doc("demo.xml")/project/project1/member
for $p2 in doc("demo.xml")/project/project2/member
where $p1//mname = "BIPIN"
and $p2//mname = "PRIYA"
return<table ID= "project">
<member>
{$p1//(mid,mname,maddress)}
{$p2//(mid,mname,maddress)}
</member>
</member>
</project1>

5. ***Find the name of the all members who have the department id same***

for $p1 in doc("demo1.xml")/project/project1/member
for $p2 in doc("demo1.xml")/project/project1/member
where $p1//did = $p2//did
and $p1//puid != $p2//puid
return<member>
<mname>{data($p1//mname)}</mname>
</member>
<member>
<mname> BI PIN</mname>
</member>
<member>
<mname> RASHI </mname>
</member>
<member>
<mname> SASHI </mname>
</member>
<member>
<mname> PRIYA </mname>
</member>

6. ***Find the project name and project id from the CSG Project1 and Project2***

for $p1 in doc("demo1.xml")//project1
for $p2 in doc("demo1.xml")//project
where $p1//topics != $p2//topics
return<table ID="project">
<pname>{data($p1//pname)}</pname>
<pid>{data($p1//pid)}</pid>
<pname>{data($p2//pname)}</pname>
<pid>{data($p2//pid)}</pid>
</table>
< table ID="project">
<pname>ABC XYZ DEF XYZ ABC</pname>
<pid> P1001 P1003 P1004 P1005 P1001</pid>
<pname> PQR</pname>
<pid> P1006</pid>
</table>

7. ***Find the details of publications where MName="Bipin" from project1 and also find the details of publication where MName="Priya" from Project2.***

for $p1 in
doc("demo1.xml")/project/project1/member
for $p2 in doc("demo1.xml")/project/project2/member
where $p1//mname = "BIPIN"
and $p2//mname = "PRIYA"
return<table ID= "project">
<publication>

```
{$p1//(puid,ptopics)}
{$p2//(puid,ptopics)}
</publication>
</table>
<table ID=”project”>
<publication>
<puid> P001</puid>
<ptopics> RRR </ptopics>
<puid> P007</puid>
<ptopics> NNN</ptopics>
</publication>
</table>
```

VI. Framework for Evaluation of Quality

The quality evaluation in multidimensional data model of conceptual level is two-fold viewpoints. Set of criteria are associate with each viewpoint, which are further defined using proposed metrics. As stated earlier, the performance (quality of results) and productiveness (time, cost, effort) of information system development may be affected by the grade of conceptual data model. In quality evaluation, the two viewpoints are (1) Designer level viewpoint and (2) User level viewpoint .The criteria like transactional/query throughput and query performance are identified by the viewpoint of designer level and the criteria like effectiveness and analyzability are identified by the viewpoint of user level.

A. Transactional/Query Throughput (QthD)

Throughput measures the number of transactions executed per second. Generally, the speed of a database in a system measured by throughput. For execution of the full set of five queries in different order, a number of query users(S) are chosen, which described in section VII Case-I and Case-II. The throughput metric is computed as the total amount of works (Sx5), converted to hours from seconds (3600 seconds per hour) and divided the total elapsed time (T_S)[Elapsed time is simply the amount of time that passes from the beginning of an event to its end.] required between the starting of first query and completion of the last one query.

$$QthD = \frac{SX5X3600}{\text{TS}} \quad (1)$$

B. Performance of the Query Execution (QEP)

Performance of query execution means the amount of time for execution of a specified query retuning with an appropriate resultant set i.e the time to make a successful one round trip.

C. Effectiveness (E)

A schema is said to be effectiveness when it represents user requirements in a natural way as well as semantic way. Measurement of effectiveness used the concept of some conceptual data model which is sufficient for exposed of some specified user requirement analysis in the system.

$$E = \frac{1}{QthD} \quad (2)$$

D. Analyzability

The *Analyzability* is the measurement of the flexibility of a user in a database model.

VII. Empirical Validations of Proposed Metrics

This portion of the article is focused on the proposed metrics and measurements empirical validation in order to prove their practical utility. The set of proposed metrics use the mechanism for guiding the grade of data models from a practical point of view that is known as the objective of empirical study. An experiment has been setup for analyzing the group of metrics and the proposed quality measurements like query throughput and query execution performance. The process of empirical validation also aims to recognize the metrics and measurement from the proposed set.

A. Experimental Settings

The desire definition of the experiment using TCSS model can be encapsulated as:

In order to analyze the set of metrics for TCSS for the purpose of evaluating if they are useful with respect of the measurement of quality of a specified data model and operability in the context of query.

Query:

To examine the scalability of proposed TCSS X-Query implementation, trying to perform an experimental evaluation using ORDER XML and PART XML database. The size of ORDER xml is 5571 KB and the size of PART xml is 1000KB respectively.

Cases:

According to the query and their types the database is organized into five basic types of queries: Selection, Retrieve, Union, Intersection and join; and the query subsets are categorized into Q1 to Q5. Trying to perform an experimental evaluation using ORDER XML database (data are taken anonymously)(Case I). The size of ORDER xml is 5571 KB(Case II).

Case I:

Trying to perform an experimental evaluation using ORDER XML database (data are taken anonymously). The size of ORDER xml is 5571 KB.

Q1: find the order status and order date from CSG order.

Q2: find the details of order where ORDERKEY=”2” and order CUSTOMERKEY= “781”.

Q3: find the details of order priority and order comment where O_ORDERKEY=”992” and “358”.

Q4: find the order STATUS which have the same CUSTOMERKEY=”317” and ORDERKEY=”998”.

Q5: find the CUSTOMERKEY and ORDERSTATUS of all orders where all the ORDERID are same.

Case II:

Trying to perform an experimental evaluation using PART XML database (data are taken anonymously). The size of PART xml is 1000KB.

Q1: find the parts name and parts Brand from CSG Part.

Q2: find the details of part where PARTKEY="3" and Part RETAILPRICE="903".

Q3:find the details of PART BRAND and PART COMMENT where P_PARTKEY="3" and "938"

Q4: find the part CONTAINER of all part which have the same PARTKEY="1960" and PARTBRAND="BRAND#33".

Q5: find the PART SIZE and PART TYPE of all parts where all the P_BRAND = "BRAND#32".

Table 2. Metrics and Measurement Value of Each Case(in Scheam level)

CASE I Query	QthD	QEP	E
Q1	10.1882	2073	0.098
Q2	9.5377	2177	0.1048
Q3	9.9475	2093	0.1005
Q4	10.050	2094.5	0.0995
Q5	9.824	2139	0.1018

CASE II Query	QthD	QEP	E
Q1	15.3191	800.5	0.0653
Q2	16.129	806.5	0.0620
Q3	15.8765	736.5	0.0630
Q4	15.852	781.5	0.0631
Q5	6.375	736	0.0611

Table 3. Collected Operation Time in ms.((TCSS-X-query)

CASE		Q1	Q2	Q3	Q4	Q5	AVG-TIME
CASE I	Comp.Time	1469	1582	1544	1473	1509	1515.4
	Eva.Time	2045	2089	2040	2082	2086	2068.4
CASE II	Comp.Time	1442	1547	1591	1460	1487	1505.4
	Eva.Time	841	840	738	731	760	782

Hypotheses: the following hypotheses are used for the experiments [28]:

- Null hypothesis (H0): Into the set of metrics and Quality Measurement as well as feasible factor of data model have no significant relationship.
- Alternate hypothesis (H1): Into the set of metrics and Quality Measurements as well as feasible factor of data model have significant relationship.

B. Experimental Steps

In order to obtain the results of the experiment it is categorized into 2 phases. In the first phase, it is checked that there are no relationship among the group of schema metrics. In the second phase, there are relationship into the set of average compile time and the average operation time, which has been evaluated to identify the group of metrics. The feasible factor of the conceptual data model has been significantly influenced by the group of metrics [28].

This is performed the independency test using non-parametric chi-square test. In both type of analysis the level of significance is set to α=0.10. Thus in both types of analysis if ρ *value* (*2tailed*) *<0.10* the null hypothesis H0 will be rejected.

Phase- I: There is no relationship between the set of schema level metrics, which are tested using non-parametric chi-square test. The result has been shown in below(spss-22):

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	20.000[a]	16	.220
Likelihood Ratio	16.094	16	.446
N of Valid Cases	5		

a. 25 cells (100.0%) have expected count less than 5. The minimum expected count is .20.

Fig.2. Phase I Chi-Square Test Query Throughput* Query Name

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	20.000[a]	16	.220
Likelihood Ratio	16.094	16	.446
N of Valid Cases	5		

a. 25 cells (100.0%) have expected count less than 5. The minimum expected count is .20.

Fig.3. Phase I Chi-Square Test QEP* Query Name

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	20.000[a]	16	.220
Likelihood Ratio	16.094	16	.446
N of Valid Cases	5		

a. 25 cells (100.0%) have expected count less than 5. The minimum expected count is .20.

Fig.4. Phase I Chi-Square Effectiveness * Query Name

The following Hypothesis is considered for the purpose.

H01: Non significant relationship among all attributes.

H11: Significant relationship among all attributes.

Reject *H01, if p*-value<0.10.

In Chi-Square Test all the obtained *p*-value are greater than α value 0.10. Hence it is significant that there is no significant relationship in all schema level metrics.

CASE I:

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of QueryThroughput is the same across categories of QueryName.	Independent-Samples Kruskal-Wallis Test	.406	Retain the null hypothesis.
2	The distribution of QEP is the same across categories of QueryName.	Independent-Samples Kruskal-Wallis Test	.406	Retain the null hypothesis.
3	The distribution of Effectiveness is the same across categories of QueryName.	Independent-Samples Kruskal-Wallis Test	.406	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Fig.5. CASE I Hypothesis Test Summary

CASE II:

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of QueryThroughput is the same across categories of QueryName.	Independent-Samples Kruskal-Wallis Test	.406	Retain the null hypothesis.
2	The distribution of QEP is the same across categories of QueryName.	Independent-Samples Kruskal-Wallis Test	.406	Retain the null hypothesis.
3	The distribution of Effectiveness is the same across categories of QueryName.	Independent-Samples Kruskal-Wallis Test	.406	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Fig.6. CASE II Hypothesis Test Summary

Phase-II: In this phase, there is a relation between proposed set of compile time and evaluation time. The set of average compile time and the average operation time, which has been evaluated to identify the existence of any significant influence of the group of metrics. The feasible factor of the conceptual data model has been significantly influenced by the group of metrics [26] .The result has been shown in below:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	20.000[a]	16	.220
Likelihood Ratio	16.094	16	.446
N of Valid Cases	5		

a. 25 cells (100.0%) have expected count less than 5. The minimum expected count is .20.

Fig.7. Phase II Chi-Square Test Compile Time*Query Name

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	20.000[a]	16	.220
Likelihood Ratio	16.094	16	.446
N of Valid Cases	5		

a. 25 cells (100.0%) have expected count less than 5. The minimum expected count is .20.

Fig.8. Phase II Chi-Square Test Evaluation Time*Query Name

Analyzing the above table, all the p-value obtained in Chi-Square Test is greater than α value 0.10 (its 0.22). it can be concluded that there exist a strong relation among compile time and evaluation time as the p-value>0.10, in each case. Hence, the proposed measure has significant influence on the Operability factor of conceptual level multidimensional data model.

VIII. CONCLUSION

In this article, a frame work of quality evaluation for conceptual level multidimensional data model has been discussed in general and for TCSS data model, in particular. The framework comprised of Direct and Indirect metrics. The set of proposed metrics use the mechanism for guiding the grade of data models from a practical point of view that is known as the objective of empirical study. An experiment has been setup for analyzing the group of metrics and the proposed quality measurements like query throughput and query execution performance. The process of empirical validation also aims to recognize the metrics and measurement from the proposed set. This article proposes a framework for quality evaluation of transactional calculus for semi-structured database system using TCSS X-Query with two different cases of data with different size.

This article also has been focused on empirical validation of the set of metrics and measurements to prove their practical utility. The several proposed metrics and operability factor of multidimensional conceptual data model significant influence is shown by empirical validation.

REFERENCES

[1] Conrad R., Scheffner D., Freytag J. C., "XML conceptual modeling using UML", 19th Intl. Conf. on Conceptual Modeling, PP: 558-574, 2000.

[2] Anirban Sarkar, "Design of Semi-structured Database System: Conceptual Model to Logical Representation",Book Titled: Designing, Engineering, and Analyzing Reliable and Efficient Software,Editors: H. Singh and K. Kaur, IGI Global Publications, USA, PP 74 – 95, 2013.

[3] McHugh J., Abiteboul S., Goldman R., Quass D., Widom J., "Lore: a database management system for semistructured data", Vol. 26 (3), PP: 54 - 66, 1997.
[4] Badia, A., "Conceptual modeling for semistructured data", 3rdInternational Conference on Web Information Systems Engineering, PP: 170 – 177, 2002.
[5] Mani M., "EReX: A Conceptual Model for XML", 2ndInternational XML Database Symposium, PP 128-142, 2004.
[6] Suresh Jagannathan,Jan Vitek,Adam Welc, Antony Hosking , A Transactional Object Calculus, Dept of Comp.sc, Purdue University,West Lafayette,IN 47906, United States.
[7] Liu H., Lu Y., Yang Q., "XML conceptual modeling with XUML", 28thInternational Conference onSoftware Engineering, PP: 973–976, 2006.
[8] Combi C., Oliboni B., "Conceptual modeling of XMLdata", ACM Symposium on Applied Computing, PP: 467– 473, 2006.
[9] Wu X., Ling T. W., Lee M. L., Dobbie G.,"Designing semistructured databases using ORA-SSmodel", 2ndInternational Conference on WebInformation Systems Engineering, Vol. 1, PP: 171 –180, 2001.
[10] Seth Gilbert and Nancy Lynch. Brewer's conjecture and the feasibility of consistent available, partition tolerant web services.SigActNews,June2002.
[11] Rita Ganguly, Rajib Kumar Chatterjee ,Anirban Sarkar. "Graph Semantic based Approach for Querying Semi-structured Database System."22nd International Conference on SEDE-2013,pp:79-84.
[12] Seth Gilbert National University of Singapore and Nancy Lynch. Brewer'sMassachusetts Institute of TechnologyPerspectiveson the CAP Theorem.
[13] Soichiro Hidaka Zhenjiang Hu Kazuhiro Inaba Hiroyuki Kato , "Bidirectionalizing Structural Recursion on Graphs",Techical Report, National Institute of Informatics, The University of Tokyo/JSPS Research Fellow, The University of Electro-Communications, August 31, 2009.
[14] Data Validation, Data Integrity, Designing Distributed Applications with Visual Studio NET,Arkady Maydanchik (2007), "Data Quality Assessment", Technics Publications, LLC
[15] Object Oriented Transaction Processing in the KeyKOS® Microkernel. William S. Frantz ,Periwinkle Computer Consulting ,16345 Englewood Ave. Los Gatos, CA USA 95032 rantz@netcom.com Charles R. Landau ,Tandem Computers Inc. 19333 Vallco Pkwy, Loc 3-22 ,Cupertino, CA USA 95014 landau_charles@tandem.com.
[16] Introduction to Object-Oriented Databases.Prof. Kazimierz Subieta, subieta@pjwstk.edu.pl,http://www.ipipan.waw.pl/~subieta Ni W., Ling T. W., "GLASS: A Graphical Query Language for Semi-structured Data", 8th International Conference onDatabase Systems for AdvancedApplications, PP 363 –370, 2003.
[17] R. K. Lomotey and R. Deters, "Datamining from document-append NoSQL," Int. J. Services Comput., vol. 2, no. 2, pp. 17–29, 2014.
[18] Braga, D., Campi, A. and Ceri, S., "XQBE (XQuery By Example): A visual interface to the standard XML querylanguage", ACM Transactions on Database Systems(TODS), Vol.30 (5), pp. 398 – 443, 2003.
[19] AnirbanSarkar, "Conceptual Level Design of Semi-structured Database System: Graph-semantic Based Approach", International Journal of Advanced Computer Science and Applications, The SAI Pubs. , New York, USA, Vol. 2, Issue 10, pp112– 121,November, 2011. [ISSN: 2156-5570(Online) &ISSN : 2158-107X(Print)].
[20] T. W. Ling. A normal form for sets of not-necessarily normalized relations. In Proceedings of the 22nd Hawaii International Conference on System Sciences, pp. 578-586. United States: IEEE Computer Society Press, 1989.
[21] T. W. Ling and L. L. Yan. NF-NR: A Practical Normal Form for Nested Relations. Journal of Systems Integration. Vol4, 1994, pp309-340.
[22] Rita Ganguly,Anirban Sarkar " Evaluations of Conceptual Models for Semi-structured Database system ".International Journal of Computer Applications.Vol 50, Issue 18, PP 5- 12,july,2012.[ISBN:973-93-80869-67-3].
[23] Rami Sellami, Sami Bhiri , and Bruno Defude, "Supporting Multi Data Stores Applications in cloud Environments." IEEE Transactions on services computing, vol-9, No-1,pp-59-71, January/February2016.
[24] O. Cur_e, R. Hecht, C. Le Duc, and M. Lamolle, "Data integration over NoSQL stores using access path based mappings," inProc. 22nd Int. Conf. Database Expert Syst. Appl., Part I, 2011, pp. 481–495.
[25] ACID vs. BASE: The Shifting pH of Database Transaction Processing,By Charles Roe ,www.dataversity net.
[26] Basili, V.R., and Weiss, D.M.,1984, " A Methodology for Collecting Valid Software Engineering Data," IEEE Transactions on Software Engineering, Vol.SE-10, No.6.,November,pp.728-738.
[27] Rita Ganguly , Anirban Sarkar," An Approach to Develop a Transactional Calculus for Semi- Structured Database System." International Journal of Computer Network and Information Security(IJCNIS), Vol.11, No.9, pp.24-39,2019. DOI:10.5815/ijcnis.2019.09.04
[28] N.G.Das., Satistical Methods, Vol.I and II ,PP :546-558,2013.

Authors' Profiles

Rita Ganguly, received the M.Tech degree from the NIT, Durgapur ,India and entitled her name as a Research Scholar (Part-time) in Computer Science department(formerly known as Computer Application department.) ,NIT, Durgapur under the supervision of Dr. Anirban Sarkar. Presently she is working as an Asst. Prof of Computer Application Department, in Dr.B.C.Roy Engineering College, Durgapur, India.

Anirban Sarkar is presently a faculty member in the Department of Computer Applications, National Institute of Technology, Durgapur, India. He received his PhD degree from National Institute of Technology, Durgapur, India in 2010. His areas of research interests are Database Systems and Software Engineering. His total numbers of publications in various international platforms are above 100. He is actively involved in collaborative research with several Institutes in India and USA and has also served in the committees of several international conferences in the area of software engineering and computer applications.

How to cite this paper: Rita Ganguly, Anirban Sarkar, "Evaluation of Quality for Semi-Structured Database System", International Journal of Computer Network and Information Security(IJCNIS), Vol.11, No.12, pp.30-39, 2019. DOI: 10.5815/ijcnis.2019.12.04

