

# Using Fuzzy Logic to Evaluate Normalization Completeness for an Improved Database Design

M. Rizwan Jameel Qureshi<sup>1</sup>

*Faculty of Computing & Information Technology of King Abdul Aziz University, Jeddah, Saudi Arabia*

*Email: anriz@hotmail.com*

Mahaboob Sharief Shaik<sup>2</sup>

*Faculty of Computing & Information Technology of King Abdul Aziz University, Jeddah, Saudi Arabia*

*Email: mshaik@kau.edu.sa*

Nayyar Iqbal<sup>3</sup>

*Department of Computer Science of Agricultural University, Faisalabad, Pakistan*

*Email: nayyar\_ciit@yahoo.com*

**Abstract**—A new approach, to measure normalization completeness for conceptual model, is introduced using quantitative fuzzy functionality in this paper. We measure the normalization completeness of the conceptual model in two steps. In the first step, different normalization techniques are analyzed up to Boyce Codd Normal Form (BCNF) to find the current normal form of the relation. In the second step, fuzzy membership values are used to scale the normal form between 0 and 1. Case studies to explain schema transformation rules and measurements. Normalization completeness is measured by considering completeness attributes, preventing attributes of the functional dependencies and total number of attributes such as if the functional dependency is non-preventing then the attributes of that functional dependency are completeness attributes. The attributes of functional dependency which prevent to go to the next normal form are called preventing attributes.

**Index Terms**—Normalization completeness, Conceptual model, Relation, Functional dependency, Total attributes, Completeness attributes, Preventing attributes

## 1. Introduction

Conceptual model describes a complete framework for a database to be designed. Conceptual model is

represented by the entity-relationship diagram or entity-relationship model that includes entities, their attributes and relationships between them<sup>[1][2]</sup>. We measure the normalization completeness for conceptual model using quantitative fuzzy functionality in two steps. Initially, we are finding the normal form of the relation by analyzing different normalization techniques up to BCNF such as checking composite attributes, partial dependencies and transitive dependencies of the relation. Normalization process<sup>[3]</sup> requires a set of dependencies to be defines for every problem. Further, we are using fuzzy membership values to scale normal form of the relation between 0 and 1.

The introduced normalization completeness determines how much the normal form is closer to the next normal form. The quality model of ISO 9126 defines functionality as a collection of attributes that engage on the existence of a set of functions and their specific properties. The functions are that satisfy stated needs which are as follow suitability, accuracy, interoperability, compliance and security<sup>[4]</sup>.

The remainder of this paper is organized as follows: Section 2 describes related work. Section 3 defines the problem statement. Section 4 presents the hypothesis. Section 5 illustrates the evaluation as a proof of concept

using a case study. Conclusion and future work are given in the final section.

## 2. Related Work

Lindlan et al. are the first to articulate a systematic framework to help understanding quality in the context of conceptual modeling [5, 6]. Previous attempts merely resulted in lists of unstructured, imprecise and often overlapping quality properties. Lindlan et al. framework is the only (framework) that contained both theory and empirical validation. The semantic quality of conceptual modeling script is difficult to evaluate directly as it is hard (and perhaps even impossible) to know reality, externalize this knowledge (which would mean building another script) and agree upon it. When evaluating semantic quality, users can only refer to their perception of reality, which is obtained through observation and internalization. The questions to find out which filter to put upon reality by our observation possibly depends on many factors such as previously acquired knowledge, perceptual psychology effects, cognitive abilities, and ontological and epistemological standpoints taken [5,6].

The research in [7] describes how to measure the semantic quality of the conceptual model using completeness. The method used to measure quantitative completeness first checks the functional dependencies. Transformation rules are applied to conceptual model and convert it into multi-graph. The concept of membership values and fuzzy hedging is used. The completeness measurement identifies the effort required for the conceptual model to transform into another conceptual model in the improved form. The quality of the conceptual model is measured using new introduced fuzzy completeness index (FCI) [8]. By considering the functional dependencies of the conceptual model the completeness of the conceptual model is measured quantitatively. The functional dependencies of the conceptual model are mapped on the TAS graph, and the

completeness of conceptual model is measured by using new FCI approach [8]. The value of the FCI determines the completeness of the conceptual model.

On schema transformation, the paper [9] presents new definitions for 'primary key', 'non-key attributes', 'key attributes' and 'functional dependency'. Schema transformation rules are also proposed in [9]. Two quality metrics are introduced namely normalization index and completeness index. Rules are applied on case studies of conceptual model to measure the normalization index and completeness index [9]. Structural complexity of a conceptual model is measured with two parameters namely modifiability and understandability [10]. Modifiability of the conceptual model is measured with effort to change. Understandability of the conceptual model is discussed by correctness and its main types are syntactic and semantic [10].

The quality of a conceptual model is divided into three types 1) syntax 2) semantic 3) pragmatic [11]. Hussain et al. [11] introduced an approach of schema transformation to improve the semantic quality of the conceptual model. The rules depend on the functional dependencies given for the conceptual model. The normal form of the conceptual model is measured up to BCNF using multiple case studies [11]. Hussain et al. [3][12] described the eliminating process of normalization. The violations while performing normalization prevent the designer to go to next form after BCNF. The normalization algorithms depend upon inclusion, multi-valued, functional and join dependencies. Removing these dependencies from given problem is a time consuming and difficult task [3][12].

The paper [11] describes the effort based completeness index for entity relationship diagram by considering the satisfying index and the effort to change for a functional dependency. The comparison of completeness index, fuzzy completeness index and effort based completeness index on different conceptual models are also shown in [13]. Two different conceptual models of the same problem can have same

completeness index but there effort based completeness index will have different value <sup>[13]</sup>.

Thalheim <sup>[14]</sup> recommends various design quality parameters for conceptual model such as flexibility, naturalness and completeness. He <sup>[14]</sup> further defines that completeness is the representation of all relevant features of the application domain. The relations are first normalized in order to obtain fuzzy relational database <sup>[15]</sup>. Fuzzy database relation has many advantages over standards database. Standard normalization depends upon the functional dependency therefore fuzzy functional dependency must be defined for fuzzy relational database normalization <sup>[15]</sup>. A model is proposed in <sup>[16]</sup> for normalization of database and functional dependencies are also introduced into the systems that needs to be considered. Inclusion dependencies are also introduced into the system <sup>[16]</sup>. The proposed model of <sup>[16]</sup> can be used to perform normalization of large database systems to remove functional dependencies.

A concept of functional dependency is introduced as rough set and relation database <sup>[17]</sup>. Functional dependency discovering algorithm is divided into two parts. In the first part, a hypothesis is defined regarding functional dependency to authenticate it against relation. In the second part, hypothesis validation is done by checking it row by row <sup>[17]</sup>. XML tree, path expressions and DTD and XML functional dependencies are described in <sup>[18]</sup> to DTD base relational schema mapping algorithm so that the semantic and structure of key can be preserved. Markus <sup>[19]</sup> describes the role of insertion, deletion and update anomalies. Semantic dependencies are the base of the code normal forms. Semantic dependencies define only one function whereas functional dependencies relates to group of attributes <sup>[19]</sup>.

The normalization theory was proposed by E. F. Codd in 1970's, and the rough set theory was introduced by Pawlak in 1982 <sup>[20]</sup>. While considering only the functional dependency then BCNF is the highest normal form in a relational database. In the relational

normalization theory, functional dependency and normal form perform the function of a kernel <sup>[20]</sup>. An automatic database normalization approach is introduced in <sup>[21]</sup>. Three structures are proposed to represent functional dependencies of the relational database that are dependency graph, direct graph matrix and dependency matrix <sup>[21]</sup>. Functional dependencies of the relation are represented by dependency graph diagram in which composite key is above the dotted line whereas other attributes of the functional dependencies are below the dotted line <sup>[21]</sup>.

A concept of normal form for XML documents is introduced in <sup>[22]</sup> so that redundancy and update anomalies can be controlled. Further, a comparison is made between XML normal form, BCNF and nested normal form. A hierarchical schema in <sup>[23]</sup> represents XML database schema and corresponding normal forms (first normal form (1NF) and second normal form (2NF)) for XML database schema. It <sup>[23]</sup> also presents the algorithm eliminating redundant schemas and normalization design algorithm up to 2NF. A concept, of functional independent normal form, is introduced in <sup>[24]</sup> considering the 'functional dependencies attributes' on the left hand side and commonly known 'determinant and attributes' on the right hand side of the functional dependency. The normal form of the database relation must be in BCNF and the conditions between the attributes of the functional dependencies must be present as follows <sup>[24]</sup>.

$$A \rightarrow B \text{ or } B \rightarrow A \text{ or } A \times B$$

Fuzzy logic <sup>[25][26]</sup> determines the membership values in numerical form that are '0' and '1'. Zero '0' means no membership and one '1' means complete membership in the condition as follows.

$$0 < x < 1$$

### 3. Problem Statement and Hypothesis

The problem statement is described as follows.

How to measure the completeness of normal form to determine that how much it is closer to the next normal form?

In the hypothesis we measure normalization completeness for the conceptual model using quantitative fuzzy functionality up to BCNF. In this research three hypothesis are considered as follows.

H1: No membership value.

H2: Partial membership value.

H3: Complete membership value.

### 4. Experiment and Analysis

Normalization completeness determines how much the normal form is closer to the next normal form. We measure normalization completeness up to BCNF.

NC = N + Fuzzy Functionality of the conceptual model ..... (1)

$$NC = N + (((\text{completeness attributes of the FD's} / \text{total attributes}) + (1 - (\text{preventing attributes of the FD's} / \text{total attributes}))) \dots (2)$$

NC is normalization completeness and N is current normal form determined by analyzing different normalization techniques discussed in <sup>[1]</sup>.

#### 4.1 Proof

In this we prove the normalization completeness for conceptual model. Fuzzy sets defined by Lotif Zadeh is given by

$$M: x \rightarrow [0, 0.01 \dots \dots 0.99, 1]$$

or

$$M: x \rightarrow [\text{no membership value, partial membership value, complete membership value}]$$

where M is fuzzy set and x describes the membership value. In which no membership value = 0, 0 < partial membership value < 1 and complete membership value = 1.

Fuzzy sets value ranges from 0 to 1.

Therefore

$$0 \leq x \leq 1 \dots \dots \dots (3)$$

suppose

$$x = (((\text{completeness attributes of the FD's} / \text{total attributes}) + (1 - (\text{preventing attributes of the FD's} / \text{total attributes}))) / 2)$$

substituting the value of x in (3)

$$0 \leq (((\text{completeness attributes of the FD's} / \text{total attributes}) + (1 - (\text{preventing attributes of the FD's} / \text{total attributes}))) / 2) \leq 1$$

Consider the total attributes of the functional dependency is n

#### Complete Membership Value

if completeness attributes = total attributes

then preventing attributes = 0

therefore

$$x = (((n / n) + (1 - (0 / n))) / 2) = (((n / n) + (1 - 0)) / 2) = ((1 + 1) / 2) = (2 / 2) = 1$$

hence it proves the complete membership is equal to 1

#### No Membership Value

if preventing attributes = total attributes

then completeness attributes = 0

therefore

$$x = (((0 / n) + (1 - (n / n))) / 2) = (((0 - 1) + (1 - 1)) / 2) = ((0 + 0) / 2) = (0 / 2) = 0$$

hence it proves the no membership is equal to 0

### Partial Membership Value

if completeness attributes  $\leq$  total attributes & preventing attributes  $\leq$  total attributes then  $x =$  partial membership value.

Hence it is proved that completeness attributes makes the normal form closer to the next normal form whereas preventing attributes decreases the completeness from the next normal form.

### Process for Finding the Normal Form

Following is the process for finding the normal form in which we analyze different normal forms techniques up to BCNF, the obtained value is assigned to N.

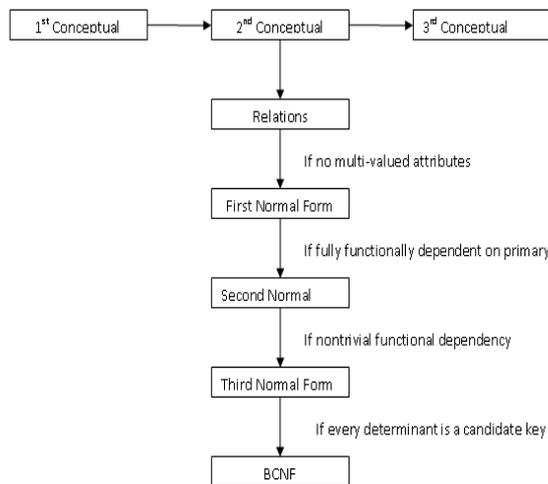


Fig. 1: Process for Finding the Normal Form

### 4.2 Case Study

The details of case study are discussed in [27].

- FD1: propertyNo, iDate  $\rightarrow$  iTime
- FD2: propertyNo, iDate  $\rightarrow$  comments
- FD3: propertyNo, iDate  $\rightarrow$  staffNo
- FD4: propertyNo, iDate  $\rightarrow$  sName
- FD5: propertyNo, iDate  $\rightarrow$  carReg
- FD6: propertyNo  $\rightarrow$  pAddress
- FD7: staffNo  $\rightarrow$  sName
- FD8: staffNo, iDate  $\rightarrow$  carReg
- FD9: carReg, iDate, iTime  $\rightarrow$  propertyNo
- FD10: carReg, iDate, iTime  $\rightarrow$  pAddress

- FD11: carReg, iDate, iTime  $\rightarrow$  comments
- FD12: carReg, iDate, iTime  $\rightarrow$  staffNo
- FD13: carReg, iDate, iTime  $\rightarrow$  sName
- FD14: staffNo, iDate, iTime  $\rightarrow$  propertyNo
- FD15: staffNo, iDate, iTime  $\rightarrow$  pAddress
- FD16: staffNo, iDate, iTime  $\rightarrow$  comments

Relation of StaffPropertyInspection are StaffPropertyInspection(propertyNo, iDate, iTime, pAddress, comments, staffNo, sName, carReg) and current normal form is 1 that means  $N=1$ .

We now find the normalization completeness: non-preventing functional dependencies are FD1, FD2, FD3, FD4, FD5, FD9, FD10, FD11, FD12, FD13, FD14, FD15 and FD16 where as preventing functional dependencies are FD6, FD7 and FD8. Completeness attributes of the FD's is '8', preventing attributes of the FD's are '6' and total attributes are equal to '8'.

$NC = N + \text{Fuzzy Functionality of the conceptual model}$   
 $NC = N + (((\text{completeness attributes of the FD's} / \text{total attributes}) + (1 - (\text{preventing attributes of the FD's} / \text{total attributes}))) / 2)$

$$\begin{aligned}
 NC &= 1 + (((8 / 8) + (1 - (6 / 8))) / 2) \\
 &= 1 + (((1) + (1 - (0.75))) / 2) \\
 &= 1 + (((1) + (0.25)) / 2) \\
 &= 1 + (1.25 / 2) \\
 &= 1 + 0.62 \\
 &= 1.62
 \end{aligned}$$

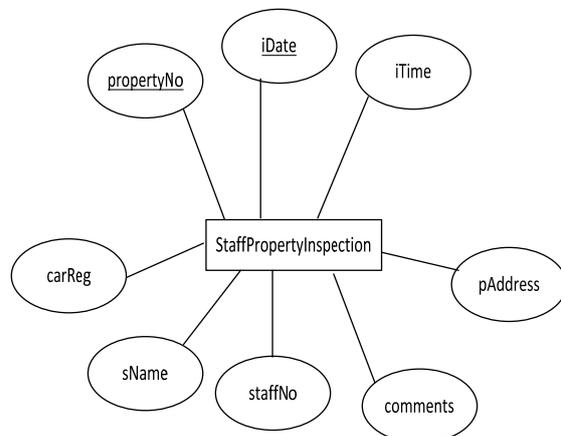


Fig. 2: Conceptual Model

Transformation of figure 2 according to [25]. According to figure 3: StaffInspection(propertyNo, iDate, iTime, comments, staffNo, sName, carReg) and current normal form is 2 therefore  $N=2$ . The normalization completeness is non-preventing functional dependencies are FD1, FD2, FD3, FD4, FD5, FD9, FD11, FD12, FD13, FD14 and FD16 where as preventing functional dependencies are FD7 and FD8. Completeness attributes of the FD's are equal to '7' and preventing attributes of the FD's are '4' and total attributes are equal to '7'.

$$\begin{aligned}
 NC &= N + \text{Fuzzy Functionality of the conceptual model} \\
 &= N + (((\text{completeness attributes of the FD's} / \text{total attributes}) + (1 - (\text{preventing attributes of the FD's} / \text{total attributes}))) / 2) \\
 &= 2 + (((7 / 7) + (1 - (4 / 7))) / 2) \\
 &= 2 + (((1) + (1 - (0.57))) / 2) \\
 &= 2 + (((1) + (0.43)) / 2) \\
 &= 2 + (1.43 / 2) \\
 &= 2 + 0.71 \\
 &= 2.71
 \end{aligned}$$

Relation of Property according to figure 3 is: (propertyNo, pAddress) and current normal form is BCNF. Relations according to figure 4 is: Inspection-(propertyNo, iDate, iTime, comments staffNo,

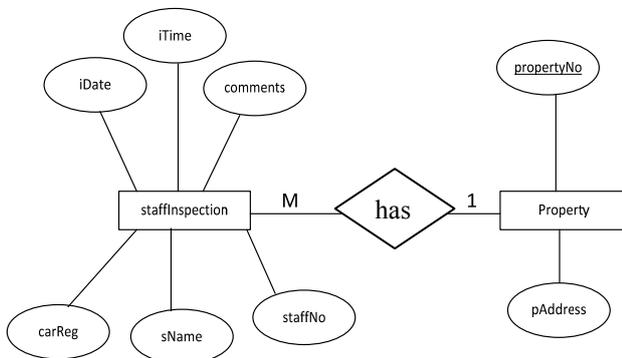


Fig. 3: Improved Conceptual Model after First Transformation

carReg) and current normal form is 3 therefore  $N=3$ . We now find the normalization completeness: non-preventing functional dependencies are FD1, FD2, FD3,

FD4, FD9, FD11, FD12, FD14, and FD16 where as Preventing functional dependency is FD8. Completeness attributes of the FD's are '6', preventing attributes of the FD's are '3' and total attributes are '6'.

$$\begin{aligned}
 NC &= N + \text{Fuzzy Functionality of the conceptual model} \\
 NC &= N + (((\text{completeness attributes of the FD's} / \text{total attributes}) + (1 - (\text{preventing attributes of the FD's} / \text{total attributes}))) / 2) \\
 &= 3 + (((6 / 6) + (1 - (3 / 6))) / 2) \\
 &= 3 + (((1) + (1 - (0.5))) / 2) \\
 &= 3 + (((1) + (0.5)) / 2) \\
 &= 3 + ((1.5) / 2) \\
 &= 3 + (1.5 / 2) \\
 &= 3 + 0.75 \\
 &= 3.75
 \end{aligned}$$

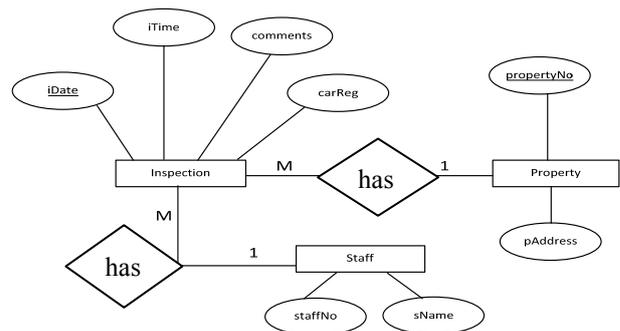


Fig. 4: Improved Conceptual Model after Second Transformation

Relation of Property according to figure 4 is Property (propertyNo, pAddress) and current normal form is BCNF. Relation of Staff [25] according to figure 4 is Staff(staffNo, sName) and current normal form is BCNF.

Table 1: Results of Case Study

Normalization completeness		
Initial Schema	After First Transformation	After Second Transformation
1.62	2.71 + 4 = 6.71	3.75 + 4 + 4 = 11.75

### 5. Conclusion

The adapted approach measures normal form for conceptual model up to Boyce Codd Normal Form (BCNF). In this paper, normalization completeness (NC)

is measured in two steps. At the first step, value of N (where N stands for normal norm) is determined by analyzing the normal form conditions. In the second step, fuzzy functionality of the conceptual model is determined that is based on hypothesis to determine the completeness of the normal form (Completeness attributes, preventing attributes of functional dependencies and total attributes are considered).

Mathematical proof for the completeness issue is based on three conditions. In first condition, if completeness attributes are equal to total attributes then preventing attributes are equal to zero. This proves that completeness membership is equal to one. In second condition, if preventing attributes are equal to total attributes then completeness attributes are zero. That means no membership is equal to zero. In third condition, if completeness attributes are less than or equal to total attributes and preventing attributes are less than or equal to total attributes, then the resultant value is the partial membership.

The normalization completeness is studied in the context of establishing a conceptual model. The study deals with two or three conceptual models. The model is improved by using transformation rules as per the literature. The model is converted to relational model and it determines normalization completeness for the conceptual model.

## References

- [1] Elmarsi, R., and Navathe, S. (2004) "Fundamentals of Database Systems", 4<sup>th</sup> Ed. Pearson Education, Inc.
- [2] Catherine, R. (2002), "Database Systems Principles, Design, and Implementation", Maxwell Macmillan International Editions.
- [3] Hussain, T., Shamail, S., and Awais, M.M., "Eliminating Normalization in Relational Database Design". In Proceeding of 7th IEEE International Multi-topic conference (INMIC 2003), December 8-9, 2003, Islamabad.
- [4] ISO/IEC 9126-1.2(1999): Information Technology – Software product quality – Part 1: Quality Model.
- [5] Ann M. and Geert P. "Evaluating quality of conceptual modeling scripts based on user perceptions", Data & Knowledge Engineering Volume 63, Issue 3, Dcember 2007, Pages 701-724, 25<sup>th</sup> international conference on conceptual modeling (ER 2006)
- [6] Keng, S., Xin, T. "Improving the quality of conceptual modeling using cognitive mapping techniques" science direct, Elsevier (2005),343-365.
- [7] Hussain, T., Shamail, S., and Awais, M.M., "A Fuzzy based Approach to Measure Completeness of an Entity-Relationship Model". In International Workshop on Quality of Information Systems, held in conjunction with 24th International conference on conceptual modeling, as LNCS, Austria, October 24-28, 2005.
- [8] Hussain, T., Shamail, S., and Awais, M.M., "Applying Fuzzy Logic to Measuring Completeness of a Conceptual Model". In Applied Mathematics and Computation, 185(2), 1078-1086, Publisher Elsevier Science, February 2007.
- [9] Hussain, T., Shamail, S., and Awais, M.M., "Schema Transformations – A Quality Perspective". In Proceedings of the International Multi-topic Conference, National University of Computer and Emerging Sciences, pp 645-649, Lahore, Pakistan, December 24-26, 2004.
- [10] Hussain, T., Shamail, S., and Awais, M.M., "On Measuring Structural Complexity of a Conceptual Model". In Proceeding of the International Conference on Software Engineering, Novosibirsk, Russia, 71-75, June 20-24, 2005.
- [11] Hussain, T., Shamail, S., and Awais, M.M., "Improving Quality in Conceptual Modeling". In the 19th Annual ACM Conference Companion on Object-Oriented Programming, Systems, Languages and Applications. (OOPSLA), Vancouver, BC, Canada, October 24-28, 2004.
- [12] Hussain, T., Awais, M.M, Shamail, S., "Multi-Valued Relationship Attributes in Extended Entity Relationship Model and Their Mapping to Relational Schema", WSEAS Transactions on Information Science and Applications 1(1), 269-273, 2004.
- [13] Hussain, T., Shamail, S., and Awais, M.M., "An Effort-based Approach to Measure Completeness of an Entity-Relationship Model". In International Workshop on Quality of Information Systems, held in conjunction with 24th International conference on conceptual modeling, as LNCS, Austria, 463-468, 2008.
- [14] Thalheim, B., "Entity-Relationship Modeling: Foundation of Database Technology". Springer Verlag, 2000.
- [15] Shirvanian, M.; Lippe, W., "Optimization of the normalization of fuzzy relational databases by using alternative methods of calculation for the Fuzzy Functional Dependency". Fuzzy Systems, 2008, (IEEE World Congress on Computational Intelligence). IEEE International Conference on 1-6 June 2008, 15–20.
- [16] Lovrencic, A.; Cubrilo, M.; Kisasondi, T., "Modelling Functional Dependencies in Databases using Mathematical Logic" Intelligent Engineering Systems, 2007. 11th International Conference on June 29 2007-July 2 2007, 307–312.
- [17] Ying Qu; Xiao-Bing Fu, "Rough Set Based Algorithm of Discovering Functional Dependencies for Relation Database". Wireless Communications, Networking and Mobile Computing, 2008. 4th International Conference on 12-14 Oct. 2008, 1–4.
- [18] Zijing Tan; JianJun Xu; Wei Wang; Baile Shi; "Storing Normalized XML Documents in Normalized Relations", Computer and Information

- Technology, 2005. The Fifth International Conference on 21-23 Sept. 2005, 123–129.
- [19] Pizka, M.; “Code Normal Forms”. Software Engineering Workshop, 2005. 29th Annual IEEE/NASA 7-7 April 2005 Page(s):97 – 108.
- [20] Z. Pawlak, Rough set, “Theoretical Aspects of Reasoning about Data”, Kluwer Academic Publisher, Dordrecht, Netherlands, 1991.
- [21] Bahmani, A.H.; Naghibzadeh, M.; Bahmani, B.; "Automatic database normalization and primary key generation" Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference on 4-7 May 2008 Page(s):000011 – 000016.
- [22] M. Arenas, L. Libkin; "A Normal Form for XML Documents." In PODS 2002 June 3-6, Page(s):85 – 96.
- [23] Y. Wu; "Normalization Design of XML Database Schema for Eliminating Redundant Schemas and Satisfying Lossless Join", International Conference on Web Intelligence 2004 Page(s):660 – 663.
- [24] Tennyson X. Chen, Sean Shuangquan Liu, Martin D. Meyer, Don Gotterbarn “An introduction to functional independency in relational database normalization”, International conference Winston Salem North Carolina USA ACM, March 23-24, 2007 Page(s):221 – 225.
- [25] L. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, IEEE Transactions on systems, Man and Cybernatics, SMC-3 (1973) 28-44.
- [26] Zadeh, Information and control (1965) 335-353.
- [27] Thomas, C., and Carolyn, B., (2007) “Database Systems A Practical Approach to Design, Implementation, and Management”, 3rd ED, Pearson Education, Inc.

**M. Rizwan Jameel Qureshi:** Assistant Professor of Faculty of Computing & Information Technology in King Abdul Aziz University, Saudi Arabia interested in software engineering and database systems.

**Mahaboob Sharief Shaik:** Lecturer, Faculty of Computing & Information Technology in King Abdul Aziz University, Saudi Arabia

**Nayyar Iqbal:** Lecturer at computer science Department at Agricultural University of Faisalabad, Pakistan major database systems and software engineering.