ISSN 2229-5518

Normalization Completeness for Conceptual Model using Quantitative Fuzzy Functionality

Nayyar Iqbal, Akmal Rehan, Dr. Khalil Ahmed

Abstract— In this new approach to measure normalization completeness for conceptual model is introduced using quantitative fuzzy functionality. We measure the normalization completeness for conceptual model in two steps. In the first step different normalization techniques are analyzed up to Boyee 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. This research is incorporating case studies to explain the normalization completeness measurement process. Normalization completeness is measured by considering completeness attributes, preventing attributes of the functional dependencies and total number of attributes. 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— Completeness attributes, Conceptual model, Functional dependency, Normalization completeness, Preventing attributes, Relation, Total attributes.

1 Introduction

Conceptual model "[1,2]" describes the complete framework of the database. Conceptual model is represented by the entity-relationship diagram or entity-relationship model which includes entities, their attributes and relationships between them. We measure the normalization completeness for conceptual model using quantitative fuzzy functionality in two steps. In the first step we find the normal form of the relation by analyzing different normalization techniques up to Boyee Codd Normal Form (BCNF) for example checking composite attributes, partial dependencies and transitive dependencies of the relation and the value obtained is the current normal form of the relation and is assigned to N. "Normalization process [3] requires a set of dependencies to be defined for every problem". In the second step we use fuzzy membership values for scaling 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 suitability, accuracy, interoperability, compliance and security [4]".

2 Related Work

In this [5] paper T. Hussain et al., described how to measure the semantic quality of the conceptual model using completeness. The method used to measure quantitative completeness first checks the functional dependencies. T. Hussain et al., apply transformation rules to conceptual model and convert it into multi-graph. In this paper [5], 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.

T. Hussain et al., [6] measured the quality of the conceptual model with new introduced fuzzy completeness index. 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 then it measures the completeness of conceptual model by using new introduced approach FCI. The value of the FCI determines the completeness of the conceptual model.

In this paper [7] "schema transformation- a quality perspective" new definitions of key, non-key attributes, key attributes and functional dependency has been discussed and also proposed schema transformation rules. Two quality metrics are introduced namely normalization index and completeness index. T. Hussain et al., apply rules on case studies of conceptual model and then measured the normalization index and completeness index.

Structural complexity [8] of a conceptual model has been measured with two parameters namely modifiability and understandability. Modifiability of the conceptual model can be measured with effort to change. Understandability of the conceptual model has been discussed by correctness, which further has two type's syntactic correctness and semantic correctness.

This paper [9] describes the quality of a conceptual model in three types 1) syntax 2) semantic 3) pragmatic. T. Hussain et al., introduced an approach schema transformation that improves the semantic quality of the conceptual model. The rules depend on the functional dependencies given for the concep-

- *Nayyar Iqbal, Lecturer Department of Computer Science, University of Agriculture, Faisalabad, Pakistan, PH-92-301-7769808. E-mail: nayyariqbal@uaf.edu.pk*
- *Akmal Rehan, Lecturer, Department of Computer Science, University of Agriculture, Faisalabad, Pakistan, PH-92-300-6676515. E-mail: akmalrehan@uaf.edu.pk*
- *Dr. Khalil Ahmed, Director, School of Computer Science, NCBA & E, Gulberg III, Lahore, Pakistan, PH-92-346-4433207. E-mail: marsnfire@yahoo.co.uk*

ISSN 2229-5518

tual model. The normal form of the conceptual model can be measured up to BCNF using case studies.

In these papers [3, 10] T. Hussain et al., explains the eliminating process of normalization and gives the detail of causes of violation of normal form up to BCNF that prevents the normal form to go into next form. 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.

Effort based [11] completeness index for entity relationship diagram can be determined by considering the satisfiability index and the effort to change for a functional dependency. T. Hussain et al., showed the comparison of completeness index and fuzzy completeness index and effort based completeness index on different conceptual models. Two different conceptual models of the same problem can have same completeness index but there effort based completeness index will have different value.

In this [12] B. Thalheim recommended following design quality parameters for conceptual model: flexibility, naturalness, minimality and completeness. "B. Thalheim describes completeness as the representation of all relevant features of the application domain".

The relations [13] are normalized first in order to obtain fuzzy relational database. Fuzzy database relation has many advantages over standards database the Shirvanian and Lippe have identified methods to remove the disadvantages. Standard normalization depends upon the functional dependency therefore fuzzy functional dependency must be defined for fuzzy relational database normalization.

In this [14] paper A. Lovrencic et al., developed a system for normalization of database and functional dependencies are also introduced into the systems that are to be considered and also introduced inclusion dependencies into the system. The introduced system can be included into larger CASE system that should draw out functional dependencies so that normalization can be done.

This paper [15] describes the concepts of functional dependency in rough set and relation database. Functional dependency discovering algorithm is divided into two parts. In the first part, hypothesis regarding functional dependency in order to authenticate it against relation is checked. In the second part hypothesis validation is done by checking it row by row.

In this [16] paper T. Zijing et al., described XML tree, path expressions and DTD and XML functional dependencies are also described. This paper introduced DTD to relational schema mapping algorithm so that the semantic and structure of key can be preserved. In this paper, it has also been proved that if DTD is normalized then relations normal form is BCNF.

In this paper [17] M. Pizka described the maintainability of code. This paper considers the concept of database normal to code. Insert, delete, update anomalies play important role in changing code. Semantic dependencies are the base of the code normal forms. Semantic dependencies define only one function whereas functional dependencies relates to group of attributes.

"The normalization [18] theory was proposed by E. F. Codd in 1970's, and the rough set theory was introduced by Pawlak in 1982." In this [19] paper it described the principles of rough sets theory and concepts of 1NF, 2NF, 3NF, BCNF. When only the functional dependencies are considered then Boyce-Codd normal form is the highest normal form in relational database. In relational normalization theory, functional dependency and normal form perform function as kernel.

This paper [20] explains the automatic database normalization approach. In this paper A. H. Bahmani et al., consider three structures to represent functional dependencies of the relational database: dependency graph, direct graph matrix and dependency matrix. 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.

This paper [21] describes the concept of normal form for XML documents. Marcelo Arenas and Leonid Libkin introduced the functional dependency concept for XML. This paper also explains the XML Normal Form (XML) so that redundancy and update anomalies can be controlled. Further the authors compare the XML normal form with Boyce Codd normal form and nested normal form.

Yonghui Wu defines "hierarchical schema [22] representing XML database schema and corresponding normal forms, first normal form (1NF) and second normal form (2NF) for XML database schema, and presents the algorithm eliminating redundant schemas and normalization design algorithm for 2NF."

In this paper [23], X. Tennyson et al., describes the functional independent normal form. The concept of functional independent normal form depends upon the functional dependencies attributes on the left hand side commonly known as determinant and attributes on the right hand side of the functional dependency. Functional independent normal form depends on the following condition as described in this paper. The normal form of the database relation must be BCNF and the following conditions between the attributes of the functional dependencies must be present "$A \rightarrow B$ or $B \rightarrow A$ or $A><B$".

Fuzzy logic [24-25] determines the membership values in numerical form which are 0, 1 and the value between them. Zero means no membership, one means complete membership and the vales in the following condition $0 < x < 1$ are called partial membership values. Antilocks braking system, bus time table, temperature control, auto-focus on a camera, predicting travel time, medical diagnoses are applications of fuzzy logic.

3 Problem Statement

To measure the completeness of normal form in order to determine how much it is closer to the next normal form.

4 Hypothesis

In this hypothesis we measure normalization completeness for 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.

ISSN 2229-5518

5 Normalization Completeness

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 for conceptual model
= N + (((CA / TA) + (1 - (PA / TA))) / 2)

Where NC is normalization completeness and N is current normal form determined by analyzing different normalization techniques discussed in [1]. CA, PA and TA stands for completeness attributes of the FD's, preventing attributes of the FD's and total attributes respectively.

5.1 Proof

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

M: x → [0, 0.01……. 0.99, 1]

or

M: x → [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 ≤ x ≤ (1)

Suppose

x = (((CA / TA) + (1 - (PA / TA))) / 2)

Substituting the value of x in euation (1)

0 ≤ (((CA / TA) + (1 - (PA / TA))) / 2) ≤ 1

Consider the total attributes of the functional dependency is n.

5.1.1 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 that complete membership is equal to 1.

5.1.2 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 that no membership is equal to 0.

5.1.3 Partial Membership Value

If completeness attributes ≤ total attributes & preventing attributes ≤ 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.

5.2 Process for Finding the Normal Form

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

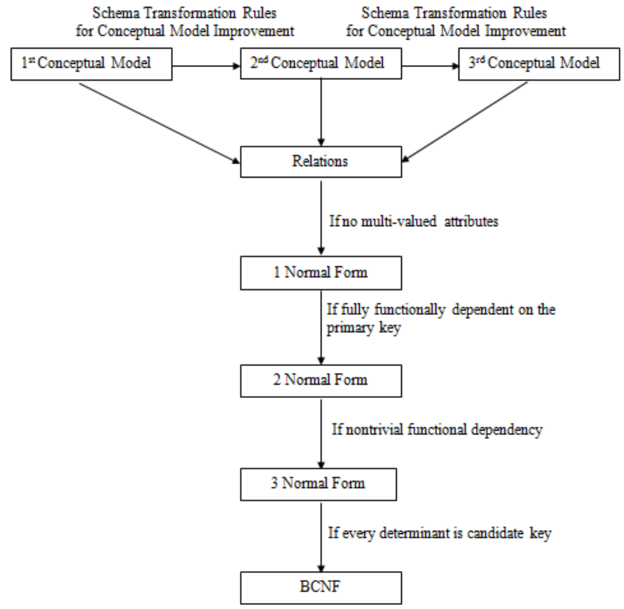

Fig. 1. Process for Finding the Normal Form

6 Case Study

This case study explains the process of normalization completeness. In this the functional dependencies defined in [26] are given in section 6.1 propertyNo, iDate is primary key.

6.1 Functional Dependencies

FD1: propertyNo, iDate → iTime
FD2: propertyNo, iDate → comments
FD3: propertyNo, iDate → staffNo
FD4: propertyNo, iDate → sName
FD5: propertyNo, iDate → carReg
FD6: propertyNo → pAddress
FD7: staffNo → sName
FD8: staffNo, iDate → carReg
FD9: carReg, iDate, iTime → propertyNo
FD10: carReg, iDate, iTime → pAddress
FD11: carReg, iDate, iTime → comments
FD12: carReg, iDate, iTime → staffNo
FD13: carReg, iDate, iTime → sName
FD14: staffNo, iDate, iTime → propertyNo
FD15: staffNo, iDate, iTime → pAddress
FD16: staffNo, iDate, iTime → comments

6.2 Conceptual Model

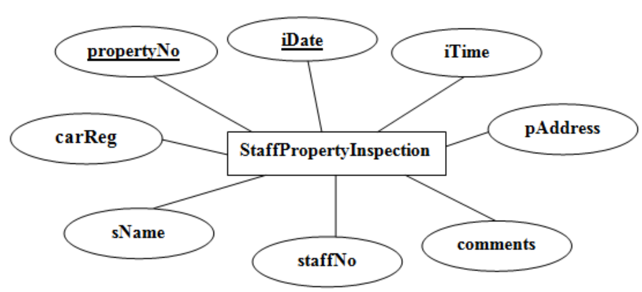

Fig. 2. Conceptual Model

International Journal of Scientific & Engineering Research, Volume 2, Issue 11, November-2011
ISSN 2229-5518

Relation of StaffPropertyIspection [26] according to fig. 2 is as follows: StaffPropertyIspection (propertyNo, iDate, iTime, pAddress, comments, staffNo, sName, carReg). Current normal form is 1 therefore 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. Preventing functional dependencies are FD6, FD7 and FD8. Completeness attributes of the FD's = 8, preventing attributes of the FD's = 6 and total attributes = 8.

NC = N + Fuzzy Functionality for conceptual model
NC = N + (((CA / TA) + (1- (PA / TA))) / 2)
= 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.

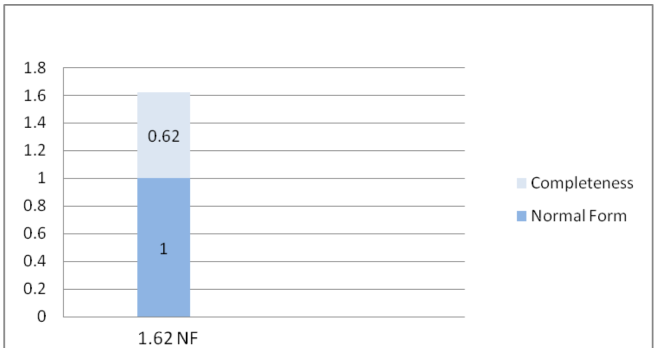

Fig. 3. Normailization Completeness for Fig.2.

6.3 Improved Conceptual Model after First Transformation

Transformation of fig. 2 according to the rules discussed in [5].

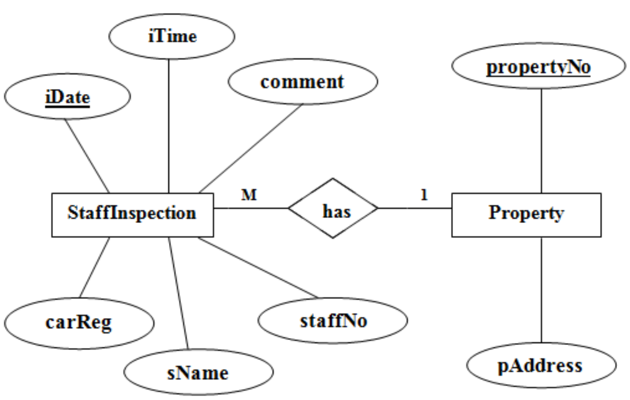

Fig. 4. Improved Conceptual Model after First Transformation

Relation of StaffInspection [26] according to fig. 4 is StaffInspection (propertryNo, iDate, iTime, comments, staffNo, sName, carReg). Current normal form is 2 therefore N=2. We now find the normalization completeness. Non-preventing functional dependencies are FD1, FD2, FD3, FD4, FD5, FD9, FD11, FD12, FD13, FD14 and FD16. Preventing functional dependencies are FD7 and FD8. Completeness attributes of the FD's = 7, preventing attributes of the FD's = 4 and total attributes = 7.

NC = N + Fuzzy Functionality for conceptual model
NC = N + (((CA / TA) + (1- (PA / TA))) / 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

Relation of Property [26] according to fig. 4 is (propertyNo, pAddress)
Current normal form is BCNF

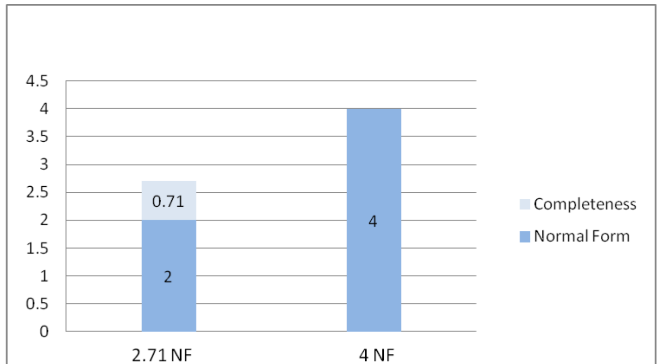

Fig. 5. Normalization Completeness for Fig. 4

6.4 Improved Conceptual Model after Second Transformation

Transformation of fig. 4 according to the rules discussed in [5].

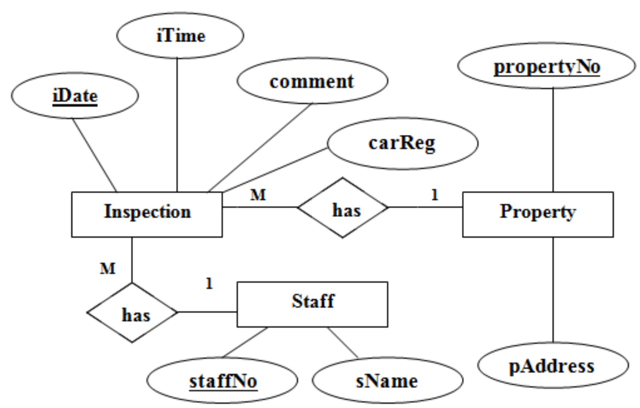

Fig. 6. Improved Conceptual Model after Second Transformation

Relation of Inspection [26] according to fig. 6 is Inspection (propertryNo, iDate, iTime, comments, staffNo, carReg). 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. Preventing functional dependency is FD8. Completeness attributes of the FD's = 6, preventing attributes of the FD's = 3 and total attributes = 6.

NC= N + Fuzzy Functionality for conceptual model
NC = N + (((CA / TA) + (1- (PA / TA))) / 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

Relation of Property [26] according to fig. 6 is Property (propertyNo, pAddress)

Current normal form is BCNF

Relation of Staff [26] according to fig. 6 is Staff (staffNo, sName)

Current normal form is BCNF

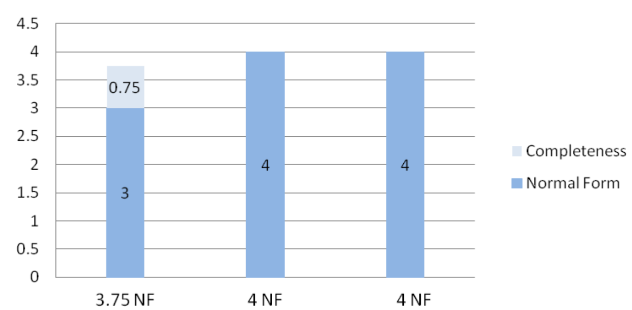

Fig. 7. Normalization Completeness for Fig. 6

6.5 Results

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

7 Conclusion

The introduced approach is used to measure normalization completeness for conceptual model. It determines how much the normal form is closer to the next normal form. This approach measure normal forms for conceptual model up to Boyee Codd Normal Form (BCNF). In this we measure Normalization Completeness (NC) in two steps. In first step, value of N (where N stands for normal form) is determined by analyzing the normal form conditions. In second step, fuzzy functionality for conceptual model is determined that is based on hypothesis it determines the completeness of the normal form, in which completeness attributes, preventing attributes of functional dependencies and total attributes are considered.

Mathematical proof for normalization completeness is also given in which total attributes are considered to be n. The proof is based on three conditions. In first condition, if completeness attributes are equal to total attributes, then preventing attributes are equal to zero. It proves completeness membership equal to one. In second condition, if preventing attributes are equal to total attributes, then completeness attributes are zero. It proves 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 applied on case study that consists of two or three conceptual models. Conceptual model is improved by using transformation rules discussed in literature. The conceptual model is converted into required relation or relations in order to determine the normalization completeness for conceptual model.

References

[1] R. Elmarsi, and S. Navathe, “Fundamentals of Database Systems”, 4th Ed., Pearson Education, Inc., 2004.

[2] R. Catherine, “Database Systems Principles, Design, and Implementation”, Maxwell Macmillan International Editions., 2002.

[3] T. Hussain, S. Shamail, and M. M Awais, “Eliminating Normalization in Relational Database Design”. In Proceeding of 7th IEEE International Multi-topic conference (INMIC 2003), Islamabad., December. 8-9, 2003.

[4] ISO/IEC 9126-1.2: “Information Technology – Software product quality – Part 1: Quality Model”., 1999.

[5] T. Hussain, S. Shamail and M. M Awais, “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.

[6] T. Hussain, S. Shamail, and M. M Awais, “Applying Fuzzy Logic to Measuring Completeness of a Conceptual Model”. In Applied Mathematics and Computation, 185(2), pp. 1078-1086, Publisher Elsevier Science., February 2007.

[7] T. Hussain, S. Shamail, and M. M 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.

[8] T. Hussain, S. Shamail and M. M Awais, “On Measuring Structural Complexity of a Conceptual Model”. In Proceeding of the International Conference on Software Engineering, Novosibirsk, Russia, pp 71-75, June 20-24., 2005.

[9] T. Hussain, S. Shamail and M. M Awais, “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.

[10] T. Hussain, M. M Awais and S. Shamail, “Multi-Valued Relationship Attributes in Extended Entity Relationship Model and Their Mapping to Relational Schema”, WSEAS Transactions on Information Science and Applications 1(1), pp. 269-273., 2004.

[11] T. Hussain, S. Shamail and M. M Awais, “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, pp 463-468., 2008.

[12] B. Thalheim, “Entity-Relationship Modeling: Foundation of Database Technology”. Springer Verlag., 2000.

[13] M. Shirvanian and W. Lippe, “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 Page(s):15 – 20., 2008.

[14] Lovrencic, A.; Cubrilo, M.; Kisasondi, T., “Modelling Functional Dependencies in Databases using Mathematical Logic” Intelligent Engineering Systems, 2007. 11th International Conference on June 29 and July 2,

ISSN 2229-5518

Page(s):307 – 312., 2007.
[15] Q. Ying and F. Xiao-Bing, “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, Page(s):1 – 4., 2008.
[16] T. Zijing, X. JianJun, W. Wei and S. Baile, “Storing Normalized XML Documents in Normalized Relations” Computer and Information Technology, 2005. The Fifth International Conference on 21-23 Sept, Page(s):123 – 129., 2005.
[17] M. Pizka, “Code Normal Forms”. Software Engineering Workshop, 2005. 29th Annual IEEE/NASA 7-7 April, Page(s):97 – 108., 2005.
[18] Z. Pawlak, Rough set, “Theoretical Aspects of Reasoning about Data”, Kluwer Academic Publisher, Dordreht, Netherlands, 1991.
[19] A. Qiusheng, W. Gaoping, Z Wenxiu, “The Study of Normal Form of Relational Database Based on Rough Sets Theory”, Granular Computing, 2007. GRC 2007. IEEE International Conference on 2-4 Nov. 2007 Page(s): 245 – 245.
[20] A.H. Bahmani, M. Naghibzadeh and B. Bahmani, "Automatic database normalization and primary key generation” Electrical and Computer Engineering, 2008. CCECE 2008. Canadian Conference on 4-7 May, Page(s):000011 – 000016., 2008.
[21] M. Arenas, L. Libkin; "A Normal Form for XML Documents.” In PODS June 3-6, Page(s):85 – 96., 2002.
[22] Y. Wu; "Normalization Design of XML Database Schema for Eliminating Redundant Schemas and Satisfying Lossless Join”, International Conference on Web Intelligence Page(s):660 – 663., 2004.
[23] X. Tennyson, Chen, L. Sean Shuangquan, D. M Martin and G. Don “An introduction to functional independency in relational database normalization”, International conference Winston Salem North Carolina USA ACM, March 23-24, Page(s):221 – 225., 2007.
[24] 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.
[25] L. Zadeh, Information and control (1965) 335-353.
[26] C. Thomas and B. Carolyn, “Database Systems A Practical Approach to Design, Implementation, and Management”, 3rd ED, Pearson Education, Inc., 2007.