I.J. Engineering and Manufacturing, **2017, 5, 65-72**
Published Online September 2017 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijem.2017.05.06

Available online at http://www.mecs-press.net/ijem

A Novel Normalization Forms for Relational Database Design throughout Matching Related Data Attribute

Youseef Alotaibi[a], Bashar Ramadan[a]

[a] *Department of Computer Science, Umm Al-Qura University, Makkah, Saudi Arabia*

Received: 20 January 2017; Accepted: 28 March 2017; Published: 08 September 2017

Abstract

One of the main goals for designing database is minimize the data redundancy. Literature shows that there is several approach have been used to minimize the data redundancy, such as converting an Entity Relational Diagram (ERD) scheme according to steps of an algorithm for ER-to-relational mapping, and applying the normalization rules. These techniques have improved the database design quality, reduced the data redundancy, and omitted from a large proportion of repetition. However, there are still a big proportion of duplicated values especially for the huge database as some attribute are related to each other and may have the same data, such as the attributes of first_name, middle_name, and family_name. These attributes cannot be called as the multivalued attribute as these values belong to variety of entities or relations. Therefore, we propose a novel normalization forms for relational database design that match the related data attribute. This proposed approach called Matching Related Data Attribute Normal Form (MRDANF). A civil registration database system is used as a case study to validate the proposed approach. The results show that using our proposed approach has the positive impact on database quality and performance as the data redundancy will be reduced.

Index Terms: Relational Database, Normalization, Multivalued Attribute, Redundancy, Related Data Attribute, Bottom-Up Design.

1. Introduction

Almost every systems or software may include a database, for example, in the bank system, hotel or airline reservation systems, health sciences system, telecommunication system, e-commerce system, etc., hence there is a need to design the database in the way ensuring that the data redundancy is minimized [1][2][16][18].

Literature shows that there is several approach have been used to minimize the data redundancy, such as converting an ERD scheme according to steps of an algorithm for ER-to-relational mapping, and applying the normalization rules [15][17]. In relational database design, normalization is the process of organizing data to

* Corresponding author.
E-mail address: yaotaibi@uqu.edu.sa

minimize redundancy [3]. It usually involves dividing the database into two or more relations or tables and defining the relationships between these tables. In general, normalization requires additional tables and some designers find this first difficult and then cumbersome [8].

These techniques have improved the database design quality, reduced the data redundancy, and omitted from a large proportion of repetition. However, there are still a big proportion of duplicated values especially for the huge database as some attribute are related to each other and may have the same data, such as the attributes of first_name, middle_name, and family_name. These attributes cannot be called as the multivalued attribute as these values belong to variety of entities or relations [11]. Therefore, we propose a novel normalization forms for relational database design that match the related data attribute. This proposed approach called Matching Related Data Attribute Normal Form (MRDANF). A civil registration database system is used as a case study to validate the proposed approach. The results show that using our proposed approach has the positive impact on database quality and performance as the data redundancy will be reduced.

The remainder of this paper is organized as follows: section 2 describes the literature review; section 3 presents our proposed normal form approach; section 4 describes the proposed framework validation with the help of a case study; and the conclusion and future research directions are presented in section 5.

2. Related Works

Literature shows that there is several approach have been used to minimize the data redundancy, such as converting an ERD scheme according to steps of an algorithm for ER-to-relational mapping, and applying the normalization rules [5][13]. In relational database design, normalization is the process of organizing data to minimize redundancy. It usually involves dividing the database into two or more relations or tables and defining the relationships between these tables. In general, normalization requires additional tables and some designers find this first difficult and then cumbersome [10][12].

Normalization was developed through 7 different standards as following.

- When using the general definitions of the first normal forms (1NF), we must be aware of multivalued attributes, composite attributes, and their combinations. It states that the domain of an attribute must include only atomic (simple, indivisible) values and that the value of any attribute in a tuple must be a single value from the domain of that attribute.
- The second normal form (2NF) was originally defined by E.F. Codd in 1971.A table that is in first normal form (1NF) must meet additional criteria if it is to qualify for second normal form. Specifically a table is in 2NF if it is in 1NF and no non-prime attribute is dependent on any proper subset of any candidate key of the table. A non-prime attribute of a table is an attribute that is not a part of any candidate key of the table. In other words, a table is in 2NF if it is in 1NF and every non-prime attribute of the table is dependent on the whole of every candidate key [6].
- A table that is in the third normal form (3NF) if the table is in 2NF and all the attributes in a table are determined only by the candidate keys of that table and not by any non-prime attribute [4].
- Boyce-Codd Normal Form (BCNF) is one of the forms of database normalization. A database table is in BCNF if there are no non-trivial functional dependencies of attributes on anything other than a superset of a candidate key [9].
- Fourth normal form (4NF) concerned with a more general type of dependency known as a multivalued dependency. A table is in 4NF if for every one of its non-trivial multivalued dependencies $X \twoheadrightarrow Y$, X is a superkey—that is, X is either a candidate key or a superset thereof [7].
- Fifth normal form (5NF)is known as project-join normal form (PJ/NF).It is a level of database normalization designed to reduce redundancy in relational databases recording multi-valued facts by isolating semantically related multiple relationships. A table is in the 5NF if every non-trivial join dependency in it is implied by the candidate keys [14].
- Sixth normal form (6NF) is a normal form for databases based on an extension of the relational algebra. In

this normal form, the relational operators, such as join, are generalized to support a natural treatment of interval data, such as sequences of dates or moments in time. A table is in 6NF if it satisfies no non-trivial join dependencies at all where, as before, a join dependency is trivial if and only if at least one of the projections (possibly U_projections) involved is taken over the set of all attributes of the table concerned.

These techniques have improved the database design quality, reduced the data redundancy, and omitted from a large proportion of repetition. However, there are still a big proportion of duplicated values especially for the huge database as some attribute are related to each other and may have the same data, such as the attributes of first_name, middle_name, and family_name. These attributes cannot be called as the multivalued attribute as these values belong to variety of entities or relations. Therefore, we propose a novel normalization forms for relational database design that match the related data attribute. This proposed approach called Matching Related Data Attribute Normal Form (MRDANF).

3. Matching Related Data Attribute Normal Form Proposed Approach (MRDANF)

Converting an ERD scheme according to steps of an algorithm for ER-to-relational mapping, and applying the normalization rules have improved the database design quality, reduced the data redundancy, and omitted from a large proportion of repetition. However, there are still a big proportion of duplicated values especially for the huge database as some attribute are related to each other and may have the same data, such as the attributes of first_name, middle_name, and family_name. These attributes cannot be called as the multivalued attribute as these values belong to variety of entities or relations. Therefore, we propose a novel normalization forms for relational database design that match the related data attribute. This proposed approach called Matching Related Data Attribute Normal Form (MRDANF).

Table 1 shows that we cannot eliminate of all duplicated values related with various entities by known normalization rules.

Table 1. Example of Duplicated Values in Database Table

Id	First Name	Middle Name	Last Name	Start_id_date
2298587871	Andrey	Henry	Smith	03/03/2015
2298738877	Henry	Andrey	Wong	15/05/2015
2189879876	James	Ramesh	Smith	03/03/2015
2178987634	Joyce	Vladimir	Henry	16/07/2015

After applying normalization rules in the relation of table 1, we note more of duplicated values especially in very large database. Therefore, we will propose a novel normalization forms for relational database design that match the related data attribute in order to eliminate the duplicated values. This proposed approach called Matching Related Data Attribute Normal Form (MRDANF).

In this proposed normalization approach, after applying all normal forms, a new type attribute called redundant attribute (RA) will be defined. This attribute will represent all redundant data in one or more attribute that belong to one or more entities. For example (first_name, middle_name, and last_name) these are redundant attribute in one entity.

In each relation, database designer must determine RA attributes from all related entities depending on the expected data as the first stage to perform our proposed MRDANF normalization approach. In this approach, a new relation (R) will be created for each RA attribute (A_1, A_2, A_n). This relation will include an attribute corresponding to all attributes (A_1, A_2, A_n) as well as the primary key attribute K—that will be as a foreign keys in primary relation instead all that attributes.

Figure 1 shows an example illustrating our proposed approach in the student relation. In this relation, the first_name, middle_name, and last_name attributes can be considered as RA attributes as they may have the same data. Thus, these RA will create a new relation called (name) by using the proposed normalization approach. This relation includes the primary key for each name as the number to defined and index the name as well as a new attribute called (name) for all available first, middle and last name without any duplication. Furthermore, the date of birth (DoB) attributes in this relation can be considered as RA attributes as they may have the same data. Thus, this RA will create a new relation called (date) by using the proposed normalization approach. This relation includes the primary key for each DoB as the number to defined and index the DoB as well as a new attribute called (DoB) for all available DoB without any duplication.

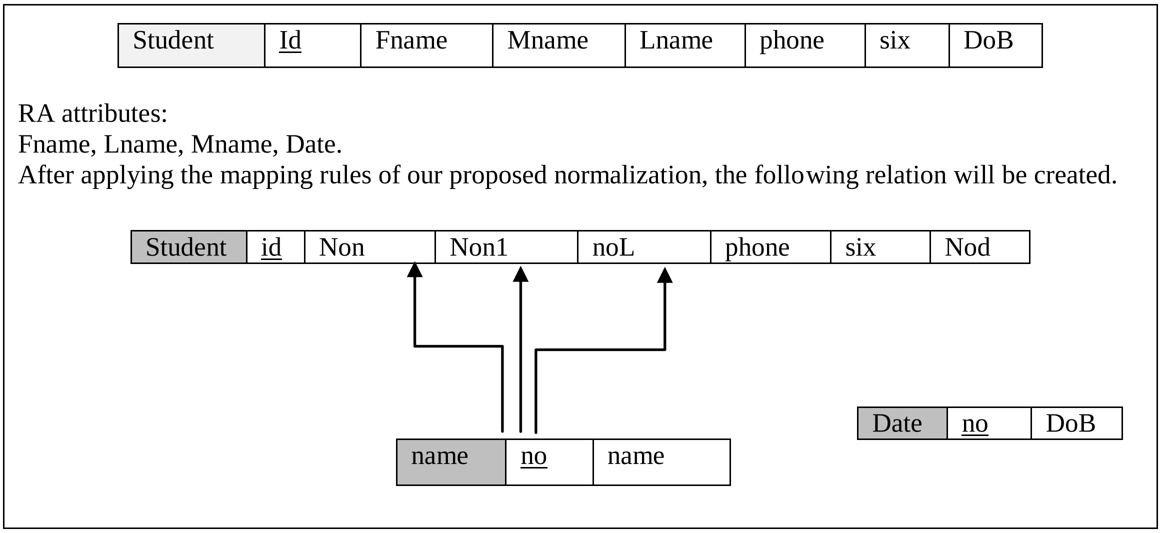

Fig.1. Example of Duplicated Values in Database Table

4. Case Study

To validate this proposed approach, the civil registration database system is used as a case study. The civil registration database system aims to keep tracking of the government records regarding the vital events (births, marriages, and deaths) of its citizens and residents. It has so many relations with redundant values such as names and dates. Two relations named called Citizen and Partner as shown in tables 2 and 3 respectively have been used as an example. The citizen relation includes social security number (Ssn) which is considered as a primary key in this relation, first name (Fname), middle name (Mname), mother name (mother_name), last name (Lname), date of birth (DoB), and issue date of ID (Iss) attributes for each citizens and residents. Moreover, the partner relation includes social security number (Ssn), social security number for partner (Ssnp) and date of marriage (DoM) attributes.

Table 2. Citizen Relation

Ssn	Fname	Mname	Mother_name	Lname	DoB	Iss
22123	Jone	Andry	Nina	Smith	02/02/99	29/09/12
22124	Henry	Jusif	Natalia	Wong	06/09/97	30/10/10
22125	Jusif	Jems	Nansi	Smith	06/09/97	30/10/10
22126	Andry	Jems	Nansi	beakar	02/02/99	29/09/12
22126	Henry	Ramis	Zenab	Wong	06/09/97	29/09/12
22127	Nina	Andry	Zenab	Beaker	07/09/94	15/05/07

Table 3. Partner Relation

Ssn	Ssnp	DoM
22140	22150	30/10/10
2290	22160	07/09/94
22111	22199	30/10/10

After applying our MRDANF mapping rules on table2 and table3, the following four relations (Figure 2) will be created with less redundant data, less size and so more efficient.

Name

no	Gname
1	Jone
2	Henry
3	Jusif
4	Andry
5	Jems
6	Ramis
7	Nina
8	Zenab
9	Nansi
10	Natalia
11	Smith
12	Wong
13	beakar

Citizen

ssn	Fn	Mn	Mon	Ln	DoB	Iss
22123	1	4	7	11	1	6
22124	2	3	10	12	2	5
22125	3	5	9	11	2	5
22126	4	5	9	13	1	6
22126	2	6	8	12	2	5
22127	7	4	8	13	4	7

Date

num	date
1	02/02/99
2	06/09/97
3	02/03/02
4	07/09/94
5	30/10/10
6	29/09/12
7	15/05/07

Partner

no	No-p	num
1	10	5
30	5	4
7	11	5

Fig.2. After Applying MRDANF Rules

5. Results

Before applying MRDANF rules assumes that the fields in citizen and partner relations have the next datatypes as shown in table 4.

If there is 6 records in citizen relation and 3 records in partner relation, the total size of these data will be 273 bytes (T1=6*(9+5+5+5+5+3+3) + 3*(9+9+3)=273 bytes).

Table 4. Data Types

Column name	Data type	Belongs to Relation	
Ssn	char(9)	Citizen	9 bytes
Fname	Nvarchar(15)	Citizen	Average 5 bytes
Mname	Nvarchar(15)	Citizen	Average 5 bytes
Mather_name	Nvarchar(15)	Citizen	Average 5 bytes
Last_name	Nvarchar(15)	Citizen	Average 5 bytes
Dob	Date	Citizen	3bytes
Iss	Date	Citizen	3bytes
Ssn	char(9)	Partner	9 bytes
Ssnp	char(9)	Partner	9 bytes
Dom	Date	Partner	3bytes

After applying MRDANF rules in the fields of citizen and partner relations, the datatypes as shown in table 5 will be used. The total size of these will be 185 bytes (T2 = 13*(1+5)+6*(9+1+1+1+1+1+1) +3*(1+1+1) +2*(1+3) = 185 bytes). Thus, we will save about 32% of the data storage as well as this percent will be increase when the database size and the number of relations' fields increase.

Therefore, the results indicate that the data storage will be saved as the data redundancy will reduced. Furthermore, the data will more efficient. In addition, using MRDANF approach will support the database designers in several issues as following.

- Security: When main table includes only keys for real data, the new gate for hacking will be closed.
- Indexing: reducing index fields and values, especially in transactional database .
- Constraints: There will be less constraints as.

Table 5. Data Types after Applying MRDANF Rules

Column name	Data type	Belongs to Relation	
No	tinyint	Name	1byte
Gname	Nvarchar(15)	Name	Average 5 bytes
Ssn	char(9)	Citizen	9bytes
Fn	tinyint	Citizen	1byte
Mn	tinyint	Citizen	1byte
Mon	tinyint	Citizen	1byte
Ln	tinyint	Citizen	1byte
DoB	tinyint	Citizen	1byte
Iss	tinyint	Partner	1byte
No	tinyint	Partner	1byte
No-p	tinyint	Partner	1byte
Num	tinyint	Partner	1byte
Num	tinyint	Date	1byte
Date	date	Date	3bytes

6. Conclusion

In this paper, we propose a novel normalization forms called MRDANF for relational database design that match the related data attribute. The civil registration database system is used as a case study to validate the proposed approach. The results show that using our proposed approach has the positive impact on database quality and performance as the data redundancy will be reduced. It may reduce the size of the data by at least 30% especially for the large database, and thus it may lead to speed up the response from the database, the paper has one limitation; we only tested our proposed approach on one database system. Thus, in the future, it could be possible to test our proposed approach with more than one database system in different business sectors.

References

[1] Alotaibi, Y. & Liu, F. (2014). "A novel secure business process modelling approach and its impact on business performance". Information Sciences. Vol. 277, pp. 375–395

[2] Alotaibi, Y. (2016). "Business process modelling challenges and solutions: a literature review". Journal of Intelligent Manufacturing, 27 (4): 701-723.

[3] Bahmani, A. H., Naghibzadeh, M., & Bahmani, B. (2008). Automatic database normalization and primary key generation. In Electrical and Computer Engineering, 2008. CCECE 2008. Canadian

Conference on (pp. 000011-000016). IEEE.
[4] Bernstein, P. A. (1976). Synthesizing third normal form relations from functional dependencies. ACM Transactions on Database Systems (TODS), 1(4), 277-298.
[5] Ch, L., & PVSS, A. (2015). Database Performance Optimization–A Rough Set Approach. International Journal of Information Technology and Computer Science (IJITCS), 7(3), 48.-53.
[6] Codd, E. F. (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM, 13 (6), 377–387
[7] Dawson, K. S., & Parker, L. P. (1988). From entity-relationship diagrams to fourth normal form: A pictorial aid to analysis. The Computer Journal, 31(3), 258-268.
[8] Diederich, J., & Milton, J. (1988). New methods and fast algorithms for database normalization. ACM Transactions on Database Systems (TODS), 13(3), 339-365.
[9] Dongare, Y. V., Dhabe, P. S. and Deshmukh, S. V. (2011) RDBNorma: "A semi-automated tool for relational database schema normalization up to third normal form", International Journal of Database Management Systems, 3 (1), 133-154.
[10] Elmasri, R., & Navathe, S. B. (2014). Fundamentals of database systems. Pearson.
[11] Fagin, R. (1977). Multivalued dependencies and a new normal form for relational databases. ACM Transactions on Database Systems (TODS), 2(3), 262-278.
[12] Fagin, R. (1981). A normal form for relational databases that is based on domains and keys. ACM Transactions on Database Systems (TODS), 6(3), 387-415.
[13] Hoang, Q., & Nguyen, V. T. (2013). Extraction of a temporal conceptual model from a relational database. International Journal of Intelligent Information and Database Systems, 7(4), 340-355.
[14] Kent, W. (1983). A simple guide to five normal forms in relational database theory. Communications of the ACM, 26(2), 120-125.
[15] Majkić, Z., & Prasad, B. (2015). Theory of sketches for database mappings. International Journal of Intelligent Information and Database Systems, 9(1), 32-53.
[16] Ougouti, N. S., Belbachir, H., & Amghar, Y. (2014). A new owl2 based approach for relational database description. International Journal of Information Technology and Computer Science (IJITCS), 7(1), 48-53.
[17] Qureshi, M., Sharif, M., & Iqbal, N. (2012). Using Fuzzy Logic to Evaluate Normalization Completeness for an Improved Database Design. International Journal of Information Technology and Computer Science (IJITCS), 4(2), 48-55.
[18] Yang, G., & Feng, J. (2012). Database Semantic Interoperability based on Information Flow Theory and Formal Concept Analysis. International Journal of Information Technology and Computer Science (IJITCS), 4(7), 33-42.

Authors' Profiles

Youseef Alotaibi is an Assistant Professor in the Department of Computer Science at Umm Al-Qura University, Saudi Arabia. He completed his PhD from the Department of Computer Science and Computer Engineering, La Trobe University in, Melbourne – Australia in 2014. He received his Master in Information Technology (Computer Network) from La Trobe University. His research interests include business process modelling, business process reengineering, information system, security, business and IT alignment, software engineering, system analysis and design, electronic commerce, and database.

Basher Ramadan is a Teaching Assistance in the Department of Computer Science at Umm Al-Qura University, Saudi Arabia. He get a postgraduate Diploma in computer network from the Department of Computer Engineering, Vladimir University in, Vladimir – Russia in 2002. He received his Bachelor in Information Technology from Vladimir University 2001. His research interests include computer network, Information security, remote control, database, and computer programing.

How to cite this paper: Youseef Alotaibi, Bashar Ramadan,"A Novel Normalization Forms for Relational Database Design throughout Matching Related Data Attribute", International Journal of Engineering and Manufacturing(IJEM), Vol.7, No.5, pp.65-72, 2017.DOI: 10.5815/ijem.2017.05.06