A Relational Data Model of Natural Langange Processing on *Nahwu* Learning

Faiz In'amurrohman[#1*], Andri Pranolo[#2*], Arief Hermawan[#1], Yana Hendriana[#2]
[#1]*Teknik Informatika, Universitas Teknologi Yogyakarta*
[#2]*Teknik Informatika, Universitas Ahmad Dahlan Yogyakarta*

Abstrak— *It is important for Muslims to study Nahwu which is became the basis for understanding the Qur'an and Hadith to avoid mistakes in interpretation. The difficulty of studying, it makes the majority of Muslims are less able to understand the Nahwu, and so needed another alternative that can facilitate Muslims to learn. This works aim to design and implementation relational database used for natural language processing application of nahwu learning. Entity relationship (ER) data model used for collecting the data from real world, furthermore it converted to relational model. Normalization process has done to ensure the relational data model produced is in 3rd Normal Form (NF). The results are ER with ten entities and relational model database which have been ten tables in 3rd NF.*

Keywords— *Natural language processing, Nahwu, Database, Entity relationship, Relational model, Normalization.*

I. INTRODUCTION

In principle, every Muslim knows that Arabic is the language of the Qur'an. Every Muslim who intends to explore the teachings of Islam, there is no other way except to be able to learn from the original source, namely the Qur'an and Sunnah. Therefore, according to the norms of Islamic law is understanding the Nahwu for peoples wanted to understand the Qur'an ruling *fard 'ain*. Each of the students at the boarding school, known that *Nahwu* usually included in the first lesson being studied and usually using Jurumiyyah book. To deepest understand Jurumiyyah, especially with memorizing is a tough task for the students, sometimes them taking a long time [1].

Based on qualitative research that has been done, the number of disciplines studied at 'Inayatullah Boarding School, Yogyakarta – Indonesia, among *Tafsir*, *Tarikh*, *Shorof*, *Fiqh*, Tawheed, and Morals makes about 83% of students are poor interested in *Nahwu*. Some of the reasons underlying the lack of interest expressed by the students among other learning methods that are less attractive, many rules contained in each chapter in the Nahwu, the material difficult and complicated to understand, as well as an explanation of the teacher is less precise and confuse students. The condition eventually make there are about 54% of students are poor to understand the material presented by the teacher Nahwu, 33% of students are fair to understand the material, and only 13% of students are good to apply the example of the problems in Nahwu. The percentage can be proved from the test results Nahwu with an average of 54.69 for students Jurumiyyah, and the average value is 54 for students' Imrithi [2].

Some works relating to the database has been done by researchers, among others are content based image retrieval from large dataset [3], design and implementation of database independent auto sequence numbers [4], comparing the performance of relational databases with post-relational databases used in Hospital Information System [5], Accesing the relational databases based on RDF View [6], fuzzy query language for relational databases [7], mapping relational database for semantic web [8]. Other works relating to the natural language (NLP) were used for hospital readmission with COPD [9], automated classification of computer-based medical device recalls [10], object oriented software development [11], "units of meaning" in medical documents [12], and to discover evidence of systems thinking [13].

This works aim to design and implementation relational database used for natural language processing application of *nahwu* learning.

II. METHOD

To design a database, conceptually and logically refers to Silberschatz *et.al.* [14], [15] which consists of the following stages:

1. Make a model of entity relationship (ER) data model. This model was built to facilitate of database design specifications. ER data model provides a means of identifying entities to be represented in the database and how those entities are related [15]. Model ER is one of several semantic data model. Semantic aspect of this model is the ability to describe the meaning of the data, that is, ER model is very useful in mapping the meanings and interactions in the real world into the concept scheme. Scheme ER model concepts illustrated with ER Diagram.
2. Reduction ER model to the relational model. The relational model contains a collection of tables which were characterized by unique names. Rows in a table representing the relationship between the set value. The concept of the relational model is illustrated by the Database Schema Diagram.
3. Normalize the tables on the relational model. Normalization is defined as the process of rearrangement of the database into a form that is normal [16]. Level of normalization consists of 1st NF, 2nd NF, 3rd NF, Boyce-Codd NF, 4th NF, dan 5th NF, and Domain/Key NF. Commonly the normalization process only done up to 3rd NF [15].

III. RESULTS AND DICSUSSION

Fig. 1 shows the relationships between the entities and their entities. On the ER diagram [15], **rectangles divided into two parts** represents entity

set, the first part contains the name of the entity set and the second part contains the names of all the attributes of the entity set. **Diamonds** represent relationship sets. **Undivided rectangles** represent the attributes of a relationship set. **Lines** link entity sets to relationship sets. **Dashed lines** link attributes of a relationship set to the relationship set. **Double lines** indicate total participation of an entity in a relationship set. Attributes that are part of the primary key are underlined. In addition, the arrow (→) indicates the cardinality of one limitation and without arrows (-) indicates many, and numbers with the format (x..y) shows the limit, x for a minimum and y for maximum participation of entities in the set of relations. Mapping cardinality consist of: one to one, one to many, many-to-one, and many to many.

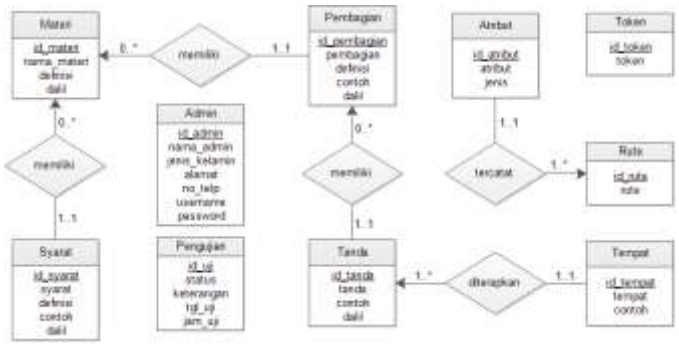

Fig. 1. Diagram of *entity relationship* model data

As an example of the relationship between the **materi entity** with the **pembagian entity** set have a mapping cardinality **one to many**, it means materi can have more than one **pembagian entity**, but much **pembagian entity** can only have one **materi**. Relations between the two entities also have a minimum and maximum limit, minimum limit indicates participation in total (1) or partial (0). Limit indicates the maximum cardinality of the relation. In this system only entity of **admin**, **pengujian**, and **token** which were did not have a relationship with other entities. All degree type of entity relationship has a binary (Fig.1).

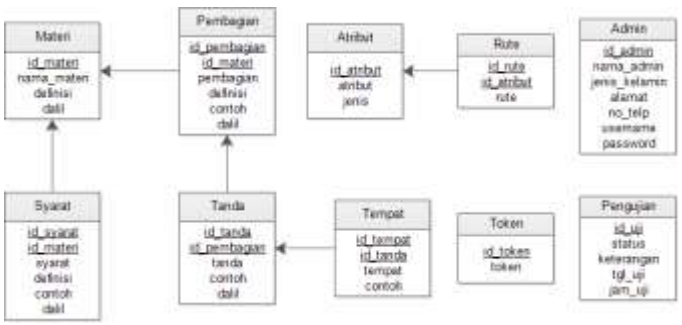

Fig. 2 Diagram of Relational model data

Fig. 2 presents the relational model that shows the result of the reduction of the ER data model (Fig. 1) to relational data model (Fig. 2). For each entity could be table in relational data model. For a binary many-to-one or one-to-many relationship set, the primary key of the entity set on the "many" side of the relationship set serves as the primary key [15]. The Foreign key created on "many" side, which it a primary key on "one" side. For example, relationship **materi entity** and **syarat entity** are "one to many", then the schema relationships become:

Materi = (id_materi, nama_materi, definisi, dalil)
Syarat = (id_syarat, id_materi, syarat, definisi, contoh, dalil)

Relational database defined as a collection of tables, each table containing rows and columns which is described as workbooks that contain multiple worksheets / spreadsheets [16], as an examples of the **Materi** table in Fig. 3.3.

id_materi	nama_materi	definisi	dalil
M-001	Kalam	Lafadz yang digunakan untuk menunjukkan makna yang bersifat murokab	kalam.jpg
M-002	Kalimah	Lafadz yang digunakan untuk menunjukkan makna yang bersifat mufrod	kalimah.jpg
M-003	Isim	Kalimah (kata) yang menunjukkan makna mandiri dan tidak disertai de	isim.jpg
M-004	Huruf	Kalimah (kata) yang menunjukkan makna apabila digabungkan dengan ka	huruf.jpg
M-005	I'rab	Perubahan akhir kalimah, baik secara perkiraan maupun secara lafadz	i'rab.jpg
M-006	Af'al	Kalimah (kata) yang menunjukkan makna mandiri dan disertai dengan p	fi'il.jpg
M-007	Mubtada	Isim yang selamanya di-rafa'-kan dan terbebas dari setiap lafadz ya	mubtada'.jpg
M-008	Khabar	Isim yang marfu' yang di-musnad-kan (disandarkan) pada mubtada kare	khabar.jpg

Fig.3 **Materi** Table

The reduction results of ER data model to relational data model obtained ten tables related to each other. Relation indicated by the lines connecting the table from one another.

In the relational database, known command of SQL (Structured Query Language) which includes DDL (Data Definition Language) and DML (Data Manipulation Language). DDL is a special language to express the database schema, and DML is a language enabled the users to access/manipulate data [15]. The following is a DDL commands to the specifications table structure of **materi**:

```
CREATE TABLE `materi` (
  `id_materi` varchar(15) NOT NULL,
  `nama_materi` varchar(25) DEFAULT NULL,
  `definisi` varchar(300) DEFAULT NULL,
  `dalil` varchar(30) DEFAULT NULL,
  PRIMARY KEY (`id_materi`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1
```

Command 1. DDL for create **materi** table

Command 1 is a DDL for create materi table. The table produced will be formed **materi** table in which **id_materi** is primary key (Fig. 4).

	Field	Type	Comment
🔑	id_materi	varchar(15) NOT NULL	
	nama_materi	varchar(25) NULL	
	definisi	varchar(300) NULL	
	dalil	varchar(30) NULL	

Fig. 4 Structure of materi table

In the relational data model (Fig. 2) appears that the **materi** table has a relationship with the **pembagian** table, so the DDL commands to create **pembagian** table with indexes, domains, relationships information, structure of physical storage, and the character set is as follows:

```
CREATE TABLE `pembagian` (
  `id_pembagian` varchar(15) NOT NULL,
  `id_materi` varchar(15) NOT NULL,
  `pembagian` varchar(25) DEFAULT NULL,
  `definisi` varchar(300) DEFAULT NULL,
  `contoh` varchar(30) DEFAULT NULL,
  `dalil` varchar(30) DEFAULT NULL,
  PRIMARY KEY (`id_pembagian`,`id_materi`),
  KEY `id_materi` (`id_materi`),
  CONSTRAINT `pembagian_ibfk_1` FOREIGN KEY (`id_materi
`) REFERENCES `materi` (`id_materi`)
) ENGINE=InnoDB DEFAULT CHARSET=latin1
```

Command 2. DDL for create **pembagian** table

Command 2 is a DDL for create **pembagian** table. The table produced will be formed **pembagian** table in which **id_pembagian** as a primary key, id_materi as a foreign key (Fig. 5).

	Field	Type	Comment
🔑	id_pembagian	varchar(15) NOT NULL	
🔑	id_materi	varchar(15) NOT NULL	
	pembagian	varchar(25) NULL	
	definisi	varchar(300) NULL	
	contoh	varchar(30) NULL	
	dalil	varchar(30) NULL	

Fig. 5 Structure of **pembagian** table

The types of DML access are: retrieval of information stored in the database (SELECT), insertion of new information into the database (INSERT), deletion of information from the database (DELETE), and modification of information stored in the database (UPDATE) [15]. Here are some examples of DML commands:

a. Insertion of new information into the database

```
INSERT INTO materi VALUES ('M-009', 'Tamyiz',
'Penjelas dari dzat yang samar', 'Tamyiz.jpg')
```

Command 3. Insertion of new information into **materi** table

Command 3 has function for insertion of new information into materi table with insertion orders are "M-009" (**kd_materi**), "Tamyiz" (**nama_materi**), "Penjelas dari dzat yang samar" (**definisi**), "Tamyiz.jpg" (**dalil**).

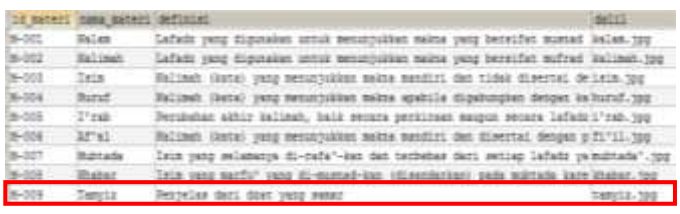

Fig 6. Result of DML insertion command (INSERT)

b. Modification of information stored in the database

```
UPDATE materi SET nama_materi='Jumlah',
definisi='Susunan kalimah', dalil='Jumlah.jpg' WHERE
id_materi='M-009'
```

Command 4. Modification of information stored in the **materi** table

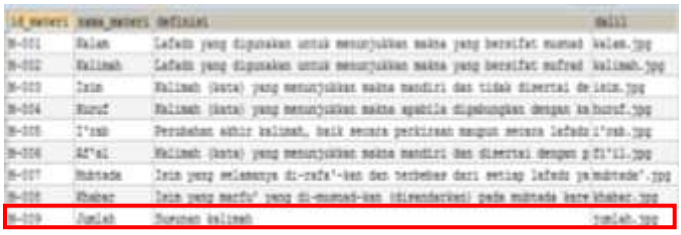

Fig 7. Modification results of **materi** table (UPDATE)

Command 4 has function for modification information stored in the **materi** table. The information which have **id_materi** is "M-009" will be modified become **nama_materi**="Jumlah", **definisi**="Susunan kalimah", and **dalil** = "Jumlah.jpg".

c. Deletion of information from Database

```
DELETE FROM materi WHERE id_materi='M-006'
```

Command 5. Deletion of information from the **materi** table

Command 5 has function for deletion of information from materi table, which **id_materi** is "M-009".

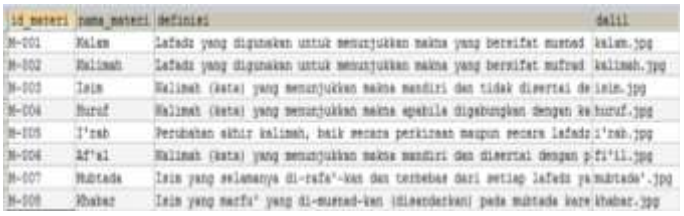

Fig 8 Deletion results of **materi** table (DELETE)

d. Retrieval of information stored in the database

```
SELECT nama_materi, definisi FROM materi
```

Command 6. Retrieval of information stored in the **materi** table

Command 6 has function for retrieval of information stored in the **materi** table. Only **nama_materi** and **definisi** column were retrieved from the **materi** table.

nama_materi	definisi
Kalam	Lafadz yang digunakan untuk menunjukkan makna yang bersifat musnad
Kalimah	Lafadz yang digunakan untuk menunjukkan makna yang bersifat mufrad
Isim	Kalimah (kata) yang menunjukkan makna mandiri dan tidak disertai de
Huruf	Kalimah (kata) yang menunjukkan makna apabila digabungkan dengan ke
I'rab	Perubahan akhir kalimah, baik secara perkiraan maupun secara lafadz
Af'al	Kalimah (kata) yang menunjukkan makna mandiri dan disertai dengan p
Mubtada	Isim yang selamanya di-rafa'-kan dan terbebas dari setiap lafadz ya
Khaber	Isim yang marfu' yang di-musnad-kan (disandarkan) pada mubtada kare

Fig. 9 Results of information retreival stored in the **materi** table (SELECT)

IV. CONCLUSION

The research results can be concluded that:

1) ER data model produces 10 entities that have a mapping cardinality one to many or many-to-one, and the reduction of ER data model to relational data model also produce 10 tables.
2) The result has been a normal form, in the 3rd NF, so it do not necessary to the normalization process.

DAFTAR PUSTAKA

[1] M. Anwar, *Ilmu Nahwu: Terjemahan Matan Al-Ajurumiyyah dan 'Imrithi Berikut Penjelasannya*. Bandung: Sinar Baru Algensindo, 1995.

[2] P. 'Inayatullah, "Laporan Rekapitulasi Hasil Ujian Madrasah," Yogyakarta, 2014.

[3] D. Kurwe, A. J. Deen, and R. Pandey, "A Hybrid Approach for content based image retrieval from large Dataset," *Int. J. Comput. Trends Technol.*, vol. 23, no. 1, pp. 16–21, 2015.

[4] K. Ray, "Design and Implementation of Database Independent Auto Sequence Numbers," *Int. J. Comput. Trends Technol.*, vol. 20, no. 2, pp. 59–61, 2015.

[5] Y. Haiyan, L. Jingsong, C. Huan, Z. Xiaoguang, T. Yu, and Y. Yibing, "Performance Evaluation of Post-Relational Database in Hospital Information Systems," in *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*, 2010, vol. 2, pp. 247–251.

[6] S.-F. Zhou, "Relational Databases Access based on RDF View," in *E-Business and E-Government (ICEE), 2010 International Conference on*, 2010, pp. 5486–5489.

[7] Y. Takahashi, "A fuzzy query language for relational databases," *Syst. Man Cybern. IEEE Trans.*, vol. 21, no. 6, pp. 1576–1579, Nov. 1991.

[8] S. Zhou, "Mapping relational database for Semantic Web," in *BioMedical Information Engineering, 2009. FBIE 2009. International Conference on Future*, 2009, pp. 521–524.

[9] A. Agarwal, S. Mulpura, R. S. Behara, and V. Tyagi, "Domain Independent Natural Language Processing -- A Case Study for Hospital Readmission with COPD," in *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on*, 2014, pp. 399–404.

[10] H. Alemzadeh, R. Hoagland, Z. Kalbarczyk, and R. K. Iyer, "Automated Classification of Computer-Based Medical Device Recalls: An Application of Natural Language Processing and Statistical Learning," in *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*, 2014, pp. 553–554.

[11] A. Tripathy and S. K. Rath, "Application of Natural Language Processing in Object Oriented Software Development," in *Recent Trends in Information Technology (ICRTIT), 2014 International Conference on*, 2014, pp. 1–7.

[12] D. Popolov and J. R. Barr, "Units of Meaning" in Medical Documents: Natural Language Processing Perspective," in *Semantic Computing (ICSC), 2014 IEEE International Conference on*, 2014, pp. 320–323.

[13] N. P. Whitehead, W. T. Scherer, and M. C. Smith, "Use of Natural Language Processing to Discover Evidence of Systems Thinking," *Syst. Journal, IEEE*, vol. PP, no. 99, pp. 1–10, 2015.

[14] A. Pranolo and S. M. Widyastuti, "Desain Basis Data Sistem Pakar untuk Identifikasi Penyakit pada Sengon (Falcataria Moluccana)," 2013.

[15] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database system concepts*, 6th ed. New York, USA: The McGraw-Hill Companies, Inc, 2011.

[16] R. Stephens, *Beginning Database Design Solutions*. Indianapolis: Wiley Publishing, Inc., 2009.