

Mechanism behind the Number of Teachers and Education Personnel: Empirical Study of Data Collection of Teacher and Education Personnel (UNTEP) in Indonesia

Sofiana Nurjanah¹, Herru Damarjati²

¹ University of Indonesia, Faculty of Computer Science, Unggulan Scholarship Beneficiary, soficitta@gmail.com

² Astra Manufacturing Polytechnic, herru.damarjati@polman.astra.ac.id

Abstract: It is important to government to anticipate the data duplication possibility in counting the number of teachers and education Personnel (TEP) that cause inaccurate data collection. Counting duplication may be occurred because of inevitable situation such as TEP can work at more than one schools without anticipation possibility that one teacher could be counted as more than one teacher. Worse, geographically it is highly possible that TEP can work for more then one school, different schools at the same districts or different schools at different districts and maybe at different provinces. Whereas, the accurate TEP data is always required by the Ministry of National Education (MONE) to make a strategic plan. In many situation inaccuracy data raise hesitations to estimate the government budget. To reduce the hesitations, MONE need the system that can detect, control and prevent the occurrence of duplication. Detection of duplication can be attempted through the similarity information attached on each individual TEP, either permanent or temporary. In Indonesia, The ministry of national education of Indonesia has unique number for teacher and education personnel (UNTEP) which well known as NUPTK in Indonesia. NUPTK also detect duplicated TEP through the similarity information attached on each individual TEP. NUPTK is generated after uniqueness ascertained and guaranteed.

Keywords : Teachers and Education Personnel (TEP), Ministry of National Education (MONE) of Indonesia, Unique Number, UNTEP, NUPTK.

1. Background

Data collection is an activity to collect some data in accordance with the type of required data. Usually the data collection based on a pre-determined data structure so that data collection can be focused and efficient. The data collection process often encounter, including the difficulty of maintaining the accuracy of the data collection. Greater amount of data collected, more difficult to be maintained the accuracy data. Data collection is to obtain a set of data which can be used as analysis materials, further analysis can be used toward organizational planning development. Inaccurate data can make the analysis also becomes inaccurate and impact to the plan implementation, such as risk and difficult to be implemented^{[1][2]}.

The government whose demographic system is not based on single ID concept yet, will perform data collection activities with single ID concept. TEP data collection should always anticipate the duplication occurrences, unless country has successfully implemented single ID concept, and TEP data are certainly a part of the Indonesian demographic system data. Duplication may be occurred because of inevitable situation such as TEP can work at more than one schools without anticipation possibility that one teacher could be counted as more than one teacher. Worse, geographically it is highly possible that TEP can work for more then one school. Inaccurate data and information

motivates MONE of Indonesia to start the new method to achieve the goal to provide completed, rational/consistency, accurate data and information. In some context, duplication can be a potential dangerous for the country.

The question is to how get accurate data and information of TEP?

2. Literature

In implementing the program, Research and Development of Education Program requires the availability of research-based data and information, and education quality standards^[1]. It includes each detail data and Information of TEP, such as historical qualification and competency data and information. Accurate data and informations can absolutely assist policy or decision making process by the ministry more effective and efficient^[2].

In general, the National Education Statistics contains data education program, particularly data related to the development of national education, such as data on schools, students, educators, and education, and infrastructure. This data was obtained by spreading the data collection instrument to each agency/institution/unit to encompass educational goals and processed data and published by the Center for Education Statistics. Data processing process of education so far has been using the concept of data base, therefore, the environment directorates, provincial education departments and district/city, need to agree on common terminology, classification systems and a consistent coding system to be used as reference in the management of educational data, hence, data collection on education this year, will conduct an inventory of data at the provincial administrative region, district, and sub-district.^[3]

Detecting name field similarity can use tools which included in several database programming languages or by creating a special algorithm. The Soundex phonetic filing system has been used for tracking genealogical data for many years. Many database systems, such as Microsoft® SQL Server, and Oracle®, have adopted the SOUNDEX as a query tool for retrieving character data based on their phonetic values^[4].

Conversion rules of soundex, the following rules are applied when calculating the SOUNDEX for a string ^[5]: (1) Keep the first letter of the string and remove all other occurrences of the following letters: a, e, h, i, o, u, w, y, (2) Assign numbers to the remaining letters (after the first) as follows: b, f, p, v = 1, c, g, j, k, q, s, x, z = 2, d, t = 3, l = 4, m, n = 5, r = 6. If two or more letters with the same number were adjacent in the original name or adjacent except any intervening h's and w's, then remove all but the first, so return the first four bytes padded with 0. For example shown as table 1^{[5] [6]}.

3. Methodology

Methodology for this research is designed to achieve the goal, there are four step as shown in figure 1.

Table 1 : Soundex Example

Function Usage	Result	Function Usage	Result
soundex('Ann');	would return 'A500'	soundex('TECH ON THE NET');	would return 'T253'
soundex('Anne');	would return 'A500'	soundex('apples');	would return 'A142'
soundex('Ansie');	would return 'A520'	soundex('apples are great');	would return 'A142'
soundex('tech on the net');	would return 'T253'	soundex('applus');	would return 'A142'

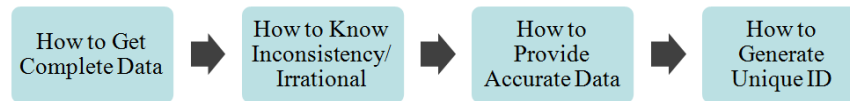


Figure 1 : Step to create the method for TEP unique ID Generation

3.1 How to Get Complete Data.

The real case of TEP data collection in Indonesia face various problem, including lack of uniformity in the system which is used by provincial office. Administrator professionalism in maintain the data and information are still less, and infrastructure readiness are also still less to perform rapid and accurate data collection in their respective regions. Availability of national consolidated data and information in the central government (MONE) may be affected by the unsolved problem yet. Data collection have been continued followed by several processes and accelerating efforts to maintain its accuracy. That effort was not fully achieved, all given constraints mentioned above is still not yet fully resolved, either by the system, human resources and leadership.

3.2 How to Know Inconsistency/Irrational

TEP data inconsistency/irrational is a common thing caused by the given time limitation for the TEP to fill their data and information (biodata). In some cases biodata filled by another person who do not know the updated data and information of TEP. Inconsistency can be corrected through verification with some limitation, such as too much the amount of inconsistency/irrational data to be verified or verification process difficult and eventually failed to be done.

3.3 How to Provide Accurate Data

Providing accurate TEP data and information requires duplication counting prevention. In duplication prevention, complete and valid data and information of TEP is a must. Duplication can also occur that can not be avoided, for example at the level of darting through the severe manipulation will be difficult to detect the validity. Geographically, in borderland, it is highly possible that TEP can work at more then one school in the different provinces. Indonesia geographic condition shows some province location are huddled to one or more than one provinces as shown in figure 2. This condition can be anticipated by including all the huddled provinces to inspect its self province. For example, Banten and West Java are huddled to Jakarta, so inspection for Jakarta must be also refered to Banten and West Java TEP data.

3.4 How to Generate Unique ID

After the data collected, complete, rational, consistent and unique data are guaranteed, the next step is generate ID for each TEP. All ID are sequantially generated. Later, ID can be used for primary key. ID can be many digit of number or maybe number and letter combination. ID may represents its self data meaning or maybe hide its self data meaning by generating encrypted number.



Figure 2 : Jakarta is huddled to Jawa Barat and Banten

4. Analysis and Design

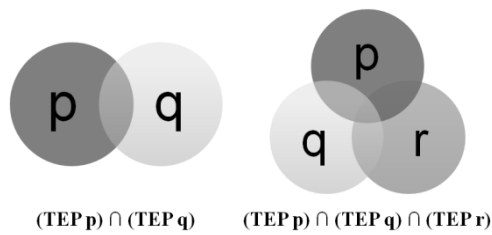


Figure 3 : The Intersection Area

All steps in methodology drive the system development. System design reflect the requirement for getting the accurate data and information of TEP. The core feature of this system is finding the intersection area member. All records in intersection area must be inspected, it may intersection from 2 variables 3 variables as shown as figure 3.

In case of TEP Data Collection, there are 7 main variables which attached on each TEP^[5] : (1) full name or name similarity, (2) birth date, (3) address, (4) school (work place), (5) mother name, (6) elementary school graduate year and (7) government employee ID. If all of variables are fully combined for inspection, system supposed to have many step inspection. All step are sorted based on their meaning and get 10 steps in highest priority, which important to do. For speed consideration, the system is decided to do only 10 highest priority inspection steps consecutively as shown in table 2.

Table 2: Thirteen inspection steps

Step	Intersection Variable			Inspection Type Action
	p	q	r	
1	Find identic records pairs			Automatic Inspection <i>Delete Record</i>
2	Employee Government ID combine with each variables			
3	Name	Birth Date	-	Manual Inspection <i>Observe one by one</i>
4	Name	Address	-	
5	Name	School	-	
6	Name	Mother's Name	-	
7	Name	Elementary School Graduate Year	-	
8	Birth Date	Address	School	
9	Birth Date	Address	Mother's Name	
10	Birth Date	Address	Elementary School Graduate Year	
11	Birth Date	School	Mother's Name	
12	Birth Date	School	Elementary School Graduate Year	
13	Birth Date	Mother's Name	Elementary School Graduate Year	

The first step will be inspected automatically because those are guaranteed as duplication pairs. Step number 2 to 6 will be inspected manually, operator will observe all the inspected records one by one. For Government Employee whose ID can be used to be an intersection variable will be inspected separately, because their low duplication probability. Logically every number generation must be pass through automatic and manual inspection. The automatic inspection can be done first, then all suspected records can be solved for number generation request manually. All process depending on completeness of the data from seven variables.

Each information of TEP are stored and used as reference variables to control on subsequent numbering. NUPTK is an Unique Identification Number, which is granted for all Teachers and Education Personnel (TEP) in Indonesia.

5. Result

Historically, number of published UNTEP as shown in table 3. In May 2008, The UNTEP system produce 2.037.058 published UNTEP. The reliable number of TEP, no incomplete, no irrational, and the most important is minimizing duplication record, as shown as figure 4.

Table 3 : Year Series Data of UNTEP

Year	Incoming Data	Publised UNTEP	Suspected as Double Counted	Teaching at more then one School
2008	2.285.579	2.037.058	248.521	88.295
2009	2.689.961	2.378.392	311.569	89.104
2010	3.149.192	2.834.556	314.636	90.553

6. Conclusion

Data and information of TEP in Indonesia have been in better condition after UNTEP system method implementation. Approach to get complete, rational, consistent and accurate data and information of TEP has been done. Overall, the impacts of using this method are distribution of TEP to all regions in Indonesia is done more proportionally by using more complete and rational data and information and also less possibility of duplication occurrences. In many situation, working with accurate data and information can make confident in making

plan and decision, such as getting precise targets, budget estimation, cost calculation, using budget efficiently and optimal, and assisting MONE in planning various welfare programs for TEP. This study requires further research for achiving the level of TEP data accuracy, which can not be determined quantitatively. The next research may also indicate what features are still not provided in this study to improve the method performance.

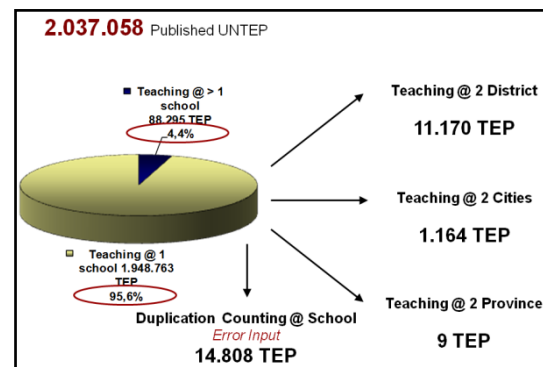


Figure 4 : TEP Data Representation, 2008

REFERENCE

- [1] Ministry of National Education Indonesia, Strategic Planning 2010-2014, (2010).
- [2] ICT Center of Ministry of National Education Indonesia, ICT Masterplan of Ministry of National Education Indonesia 2009-2013, (2009)
- [3] Human Resource and Development of Ministry of National Education Indonesia, Education Data Collection Guideline, (2010).
- [4] Zizhong Fan, Matching Character Variables by Sound: A closer look at SOUNDEX function and Sounds-Like Operator (=*), (2002).
- [5] Oracle FAQ, SOUNDEX, <http://www.orafaq.com/wiki/SOUNDEX>, (May 10, 2011:06.00).
- [6] Tech On The Net, Oracle/PLSQL: Soundex Function, <http://www.techonthenet.com> (May 10, 2011:08.50)
- [7] GoogleMap, DKI Jakarta – Indonesia, May 10, 2011:22.00)
- [8] Nurjanah, Sofiana., How to Provide Accurate Data in Teacher and Education Personnel Data Collection: an Application Development Process, (2011).