SPREADSHEET AS A MEANS TO SUPPORT WORKING TIME SCHEDULING – DATA INTEGRITY IN A SPREADSHEET TOOL

Bartosz SZCZĘŚNIAK
Silesian University of Technology, Faculty of Organization and Management, Institute of Economics and Informatics; bartosz.szczesniak@polsl.pl, ORCID: 0000-0002-9683-4629

Abstract: Spreadsheets are commonly used in nearly every organization. The tools created by their means may be used to support a multitude of diversified information processes handled in businesses. Such tools can be built in accordance with the concept of ‘Microtools Based on the Relational Data’. One of the problems that must be addressed while such a tool is being created is to ensure the integrity of the data stored in the tool. The article addresses a selection of solutions proposed and applied in this respect in a tool supporting a process in which the working time of maintenance department employees is scheduled.

Keywords: business process improvement, database, relational data model, information process, MiRel.

1. Introduction

Spreadsheets are undoubtedly among the most popular universally applied computer programs used in business on a daily basis. The range of their potential applications is very extensive, covering such spheres as management in finance, accounting, controlling (Winston, 2011; Jackson, and Staumont, 2004; Carlberg, 2009), quality management (Knight, 2006; Szczęśniak, 2010), or management in sales and marketing (Wilczewski, and Wrzód 2008; Winston, 2011). Spreadsheets also provide a very good platform for independently building original IT tools supporting processes unique to individual enterprises. Such tools can be created spontaneously, without having explicitly defined the structure of the data contained in the given tool in advance, or in a more organized manner, where the process of creating tools must be preceded by analyzing the area to be supported with enough precision to enable accurate determination of the data structure on which the tool will be based. What one can observe in publications of numerous authors is that their spreadsheet related concepts interpenetrate with those concerning relational databases (Cunha et al., 2009; Witkowski et al., 2003; Witkowski et al., 2005). It has also been proposed that a spreadsheet can be used as
a relational database engine (Tyszkiewicz, 2010; Sroka et al. 2015). The necessity to identify in advance the relational model of the data to be stored in the tool, as well as the organized approach to developing the model are the essential features of the concept of ‘Microtools Based on the Relational Data Model’ (Szczęśniak, 2016, 2017, 2018a, 2018b). An example of a tool created in accordance with this concept is a solution that supports the work time scheduling of maintenance department employees. The general concept behind the scheduling process and the proposed solution were presented in an earlier publication (Zasadzień et al., 2017). This article discusses the approach used in the aforementioned tool to ensure the integrity of the data entered into it.

2. Data structure applied in a tool developed to support scheduling operations

An analysis of a solution developed to schedule the work times of maintenance department employees has made it possible to identify entities that should be included in the tool. These entities are “Year”, “Week”, “Day”, “Task”, “Error”, “Employee”, “Employee in week”, “Task in week”, “Employee’s task on weekday”. The structure of these entities, along with the interrelations between them and the identified attributes is shown in Figure 1, using a diagram conforming with the CASE METHOD approach.

![Figure 1. Structure of entities in the tool discussed.](image)

In this tool, entities such as “Year”, “Week”, “Day”, “Task”, “Error”, “Employee” and “Task in week” have been presented as classical tables having the same layout as a standard relational database. The table representing the “Task in week” entity has been entered into worksheet Z in rows 1-2000. One-column tables representing the remaining entities have all been placed in worksheet D in rows 1-100. The layout of the aforementioned tables is presented in Figure 2.
Spreadsheet as a means to support working time scheduling

Figure 2. Representation of entities in worksheets Z and D.

Entities “Employee in week” and “Employee’s task on weekday” have been jointly presented in an extended table, combining both type I and type II of entity representation (Szczęśniak, 2017), and placed in worksheet H in rows 2-4164. The table’s layout is shown in Figure 3. In this figure, the letters “A” and “B” mark the cells which contain attribute values for a sample single instance of entities, i.e. “Employee in week” and “Employee’s task on weekday” respectively.

Figure 3. Representation of entities in worksheet H.

3. Solutions ensuring data integrity

In order to ensure data integrity, the tool must feature solutions to certain problems. One of these is providing conformity between values of foreign keys and the entered values of the corresponding primary keys. What has been proposed for this purpose is to make use of the Data Validation mechanism in such a way as to enable the values of respective attributes to be selected from drop-down lists. The Name Manager mechanism has also been used in this respect, so that the sources of data for individual lists are named worksheet ranges that have previously been adequately defined. In range E5:CJ4164 of worksheet H, uniform data validation settings have been applied for all cells, and the reference to the corresponding named cell range depends on the name of an attribute in the given row. Since the values of attributes “Time_P” and “Time_R” are not subject to the key conformity limitation, the drop-down lists used in this range are not restrictive, but only supportive, and in the case of both of these
attributes, the source for the list is an empty range named “d_Any”. Table 1 lists the named ranges defined in the tool, along with the ranges in which they have been used.

Table 1.
Source ranges for drop-down lists of values

<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Target Worksheet</th>
<th>Target Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>d_Error_no</td>
<td>=OFFSET(D!$E$2;0;0;COUNTA(D!$E$2:$E$100);1)</td>
<td>H</td>
<td>E5:CJ4164</td>
</tr>
<tr>
<td>d_Task_no</td>
<td>=OFFSET(D!$D$2;0;0;COUNTA(D!$D$2:$D$100);1)</td>
<td>Z</td>
<td>D2:D2000</td>
</tr>
<tr>
<td>d_Year</td>
<td>=OFFSET(D!$B$2;0;0;COUNTA(D!$B$2:$B$100);1)</td>
<td>H</td>
<td>E5:CJ4164</td>
</tr>
<tr>
<td>d_Employee</td>
<td>=OFFSET(D!$A$2;0;0;COUNTA(D!$A$2:$A$100);1)</td>
<td>H</td>
<td>C5:C4164</td>
</tr>
<tr>
<td>d_Week</td>
<td>=OFFSET(D!$A$2;0;0;COUNTA(D!$A$2:$A$100);1)</td>
<td>Z</td>
<td>B2:B2000</td>
</tr>
<tr>
<td>d_Any</td>
<td>=OFFSET(D!$A$2;0;0;COUNTA(D!$A$2:$A$100);1)</td>
<td>H</td>
<td>A5:A4164</td>
</tr>
</tbody>
</table>

Beyond the data validation mechanism that has been applied in worksheet H, this contains a solution which identifies cases where a “Year-Week-Task_no” combination that does not appear in worksheet Z, has been introduced. For this purpose, auxiliary column C20 has been added to worksheet Z, while auxiliary columns C1 and C2, and table T1 (whose cells correspond to the cells of range E5:CJ4164) have been added to worksheet H. The auxiliary column entered in worksheet Z and its formula are shown in Figure 4.

Figure 4. Validation of the “Year-Week-Task_no” combination – auxiliary column in worksheet Z.

All formulas presented in the article are uniform in form across the entire range to which they have been assigned, and they are presented for the top left cell of that range. For example, the F4A formula shown in Figure 4 corresponds to cell G2. Its form in all the other cells in column C20 results from copying the formula from cell G2 to all the other cells in range C2:C2000.

The additional cells in worksheet H and the respective formulas they contain have been provided in Figure 5. The values of “1” in column C2 mean that the values of attributes “Year” and “Week” in the current row should be highlighted. The values of “1” in table T1 imply that the items which should be highlighted are the corresponding values of attribute “Task_no”. The highlighting is accomplished using conditional formatting.
Spreadsheet as a means to support working time scheduling

Figure 5. Validation of the “Year-Week-Task_no” combination – auxiliary cells and formulas in worksheet H.

The second problem in the field of data integrity is verifying whether values have been entered for all the required attributes. For the “Task in week” entity, none of the attributes is optional, which is why entering a value for any attribute in the table in worksheet Z requires that all other attributes are also assigned values. The mechanism for identifying cells with missing values is shown in Figure 6.

Figure 6. Identification of missing attribute values – worksheet Z.

If the values of “1” appear in auxiliary table T20, this means that, although the corresponding cells in range B2:E2000 should contain the values entered for the corresponding attributes, they actually do not have these values. These cells are highlighted using conditional formatting.

In the case of entities represented in worksheet H, the first of the solutions used to identify the missing values is in Figure 7. This solution pertains to a situation where a value of any of attributes “Task_no”, “Time_P”, “Time_R” and “Error_no” appears in range E5:CJ4164. In such an event, it is necessary to enter the values of attributes “Year”, “Week”, “Employee”, “Time_P” and “Task_no” for the current instance of the “Employee’s task on weekday” entity and the values of attributes “Time” and “Overtime” for the associated instance of the “Employee in week” entity. As a part of the solution in question, 3 columns and 3 auxiliary tables have been added to worksheet H. The “R/O” values in column “C4” determine the optionality of corresponding attributes “Task_no”, “Time_P”, “Time_R” and “Error_no”, and on account of their fixed structure in the tool, they have been entered manually. The values
in other auxiliary cells are determined using the formulas provided in Figure 7. The numbers “1” in auxiliary tables T2, T3 and T4 mean that the corresponding cells, in which individual attributes are entered, remain empty even though they should contain values previously entered, and they consequently become highlighted. Table T2 is associated with the cells in range A5:C4164, and it identifies the missing values of attributes “Year”, “Week” and “Employee”. Table T3 is associated with the cells in range C15:CL4164, i.e. it is linked with attributes “Time” and “Overtime”. Table T4 is associated with the cells in range E5:CJ4164 “Task_no”, and it is linked with attributes “Task_no”, “Time_P”, “Time_R” and “Error_no”.

![Figure 7](image)

**Figure 7. Identification of missing attribute values – worksheet H, solution 1.**

The second solution introduced in worksheet H applies to a situation where a value of any of attributes “Year”, “Week” or “Employee” has been entered in the table in range A5:C4164, or a value of attributes “Time” or “Overtime” has been entered in range CK5:CL4164. Where this is the case, one must enter all the remaining attributes for the given instance of entity “Employee in week”. The mechanism which identifies the missing values in this case is shown in Figure 8.

![Figure 8](image)

**Figure 8. Identification of missing attribute values – worksheet H, solution 2.**
The numbers “1” appearing in the cells of auxiliary tables T5 and T6 imply that, in the corresponding cells of ranges A5:C4164 and CK5:CL4164, respectively, there is a required attribute value missing, and so they will be highlighted by way of conditional formatting.

The third problem considered in terms of ensuring data integrity in the tool is how to provide for the uniqueness of primary keys. The necessity to identify cases of primary key values that appear repeatedly pertains to each of worksheets D, Z and H. The simplest case is that of worksheet D, where it is required to highlight cells containing repeated single values, separately within each of the columns in use, i.e. “Employee”, “Year”, “Week_no”, “Task_no”, “Error_no” and “Day_no”. For this purpose, one can directly apply the conditional formatting mechanism with an option of duplicate value formatting. What is required in worksheet Z is that the cells containing repeated combinations of values of the “Year-Week-Task_no” attribute are highlighted. To address this issue, three auxiliary columns have been added. Their arrangement, along with the applied formulas is shown in Figure 9.

![Figure 9. Identification of repeated Year-Week_Task_no pairs.](image)

In worksheet H, in the table that represents the “Employee in week” and “Employee’s task on weekday” entities, the problem with the uniqueness of primary keys is more complex. In this table, the given combination of values of attributes “Year”, “Week” and “Employee” can only occur in one group of four adjacent rows containing attribute values associated with one instance of the “Employee’s task on weekday” entity and one instance of the “Employee in week” entity. It is necessary to use two mechanisms in this case. One of these checks whether all values of the said attributes appearing in four adjacent rows are identical. The second mechanism checks if the combination of values of the attributes in question appearing in one group of four adjacent rows does not repeat itself in any other row in the table. Under both mechanisms, auxiliary columns have been added to worksheet H. The columns added as part of the first mechanism along with the formulas applied in them is shown in Figure 10. The values of “1” which appear in column C10 indicate a situation where, in the given group
of four adjacent rows, the values of attributes “Year”, “Week” and “Employee” are not identical in each row. The values one will find in column C10 provide grounds for using conditional formatting to highlight cells containing invalid values. All cells containing the values of the attributes in question in the given group of four adjacent rows are highlighted as a whole.

![Figure 10. Mechanism validating the values of attributes “Year”, “Week” and “Employee” in a group of four adjacent rows.](image)

The columns added under the second mechanism along with their formulas are shown in Figure 11. The values of “1” appearing in column C13 imply that the “Year-Week-Employee” combination in the given row can also be found in at least one row that does not belong to the given group of four adjacent rows.

![Figure 11. Mechanism verifying the uniqueness of the values of attributes “Year”, “Week” and “Employee”.](image)
The values of “1” can be found in all rows containing the combinations of these attributes which repeat themselves in an inappropriate manner. With reference to the values in column C13, cells that contain invalid values of attributes “Year”, “Week” and “Employee” are highlighted using conditional formatting.

In the table in worksheet H, it is also necessary to verify the uniqueness of the “Week-Year-Employee-Day-Task_no” attribute combination. For this purpose, one auxiliary table T7 has been added to the worksheet, as shown in Figure 12, along with the formula it contains.

![Figure 12. Mechanism verifying the uniqueness of the combination of attributes Year, Week, Employee, Day and Task_no.](image)

On account of the table’s structure, as well as the fact that mechanisms intended to verify the uniqueness of the “Year-Week_Employee” combination have already been applied, in this case, it is only necessary to check that the value of attribute “Task_no” does not repeat itself in the given row within one day. The cells in table T7 correspond to individual cells in range E5:CJ4164. If the value of “1” appears in the table, the values of attribute “Task_no” in the corresponding cells repeat themselves within the same day, and should, therefore, be highlighted as cells containing invalid values. As in the solutions previously addressed, so in the case in question, the highlighting is done by conditional formatting.

4. Conclusions

With regard to a tool developed in accordance with the concept of Microtools Based on the Relational Data Model, specific solutions that support the integrity of the data stored in the tool have been proposed and discussed in the paper. The aforementioned solutions make it possible to easily identify and eliminate problems related to entering values of all the required attributes for the given instance of entities, the conformity between values of foreign keys and primary keys in source tables, as well as the uniqueness of values of primary keys. The verification mechanisms proposed in the article make it possible to ensure data integrity in the event that...
entities are represented as classical tables, such as those used in standard relational databases, as well as where they are represented in the form of extended tables combining type I and type II of entity representation assumed under the MiRel concept. In the second case, the solutions applied must be more extensive and complex. All the mechanisms addressed by the author have been developed using a relatively small set of standard spreadsheet features and components, including such functions as OFFSET(), COUNTA(), IF(), CONCATENATE(), AND(), COUNTIF(), SUM(), MOD(), QUOTIENT(), ROW(), COUNTIFS(), COLUMN(). Some other spreadsheet features used in the process are Data Validation, Name Manager and Conditional Formatting. All of the assumed functionality has been achieved without having to create any code using any programming language.

References

Spreadsheet as a means to support working time scheduling


