A VISUALIZATION ARTEFACT TO SUPPORT LEARNING IN DATABASES

Imelda Smit[1] & Romeo Botes[2]
North-West University
imelda.smit@nwu.ac.za[1], romeo.botes@nwu.ac.za[2]

ABSTRACT– This paper focuses on how the concept of visualization can assist learners in their endeavour to understand, model and practically implement databases. From the perspective of an honours learner working as a student assistant in databases, as well as being a learner in that same subject modules only recently; the difficulties learners experience in this class initiated the identification of a problem. The main obstacles include understanding, interpreting and compiling Entity Relationship Diagrams that needs to be implemented as database tables before the construction of Structured Query Language queries can commence. In assisting learners to overcome the struggles mentioned, the idea of creating an artefact that will help learners to visualize the link between an Entity Relationship Diagram and the queries to be built, presented itself. The main objective of creating the suggested visualization artefact in this context is to facilitate the mapping between the Entity Relationship Diagram and the queries to be built. Understanding how to compile a logical Entity Relationship Diagram, as well as its physical counterpart, the Database Schema, are both important modelling tools to include in a third year computer science learner's skills list when they finish their degree course and start working. Likewise, the implementation of the model in a database environment and the subsequent manipulation of the data are valuable skills. The suggested artefact intends enhancing learners' understanding of the analysis, design, development, and support of databases through visualization. Since the development of an artefact is only a first step towards assisting learners, especially those from disadvantaged communities, potential further studies are scrutinized.

Keywords: Visualization artefact; Entity Relationship Diagram (ERD); Structured Query Language (SQL)

1 INTRODUCTION

Confusion regarding how to compile an Entity Relationship Diagram (ERD), distinguish between its logical and physical foci, as well as understanding how to use the compiled ERD representing a database scenario to assist in building Structured Query Language (SQL) queries are common in the early days of making sense of databases. Experience as a recent third year learner doing the database subject modules, as well as being a student assistant for these modules, made it easy to identify with the learners' struggle. The idea of the creation of an artefact came to mind, and specifically a visualization artefact.

The primary goal of this study is to design and develop a possible solution to help learners to visualize how an ERD and SQL queries interact with one another. The study was conducted at the Vaal Triangle Campus of the North-West University where the database subject modules are offered as part of the Information Technology course.
In the subsequent sections, the topics addressed are; a literature review discussing concepts including visualization and the factors to consider in developing visualization artefacts, the ERD, SQL and the software utilized; the methodology employed, namely Design Science Research (DSR), including the steps followed and the artefact's description; and finally the conclusion of the paper.

2 LITERATURE REVIEW

Visualization and the factors to be utilized after artefact development, ER diagrams, and SQL are scrutinised subsequently to facilitate clarity.

2.1 Visualization and factors to consider in developing visualization artefacts

The concept of visualization is explained as the use of visual images and thinking in a visual manner to understand and grasp a concept based on difficult information (Ursyn, 2014). Focusing on a

specific topic in the form of pictures and seeing it as a representation of difficult concepts, can help one to understand it better through a visual representation.

Factors to consider when building a visualization artefact, include flow of events between the ERD and the SQL queries being built, interaction between the user – the learner and the artefact, and the presentation of the different components relating to the ERD pertaining to a query that is built (Liang, 2010). The factor of *flow* is seen as the time of interaction with visual representations and how visually attractive these are made to reflect the order of events (Agosti, Foner, Muller, & Santucci, 2013). Liang (2010) describes the second factor, *interaction*, as an action being executed by an artefact with a subsequent reaction needed from the user to be able to progress. The third factor of *visualization* is described by Ursyn (2014) as using visual techniques and images created to facilitate the understanding of information. Fishman and O'Conner-Divelbiss (2000) adds that a good layout, easy interaction and a user friendly interface of a visualization artefact will contribute to the effectiveness and successfulness of the end product. Phillips, Stephen, and Norris (2010) combine the three factors mentioned above by stating that the artefact should use visualization to break down the whole into its components and then show how different components fit together to create the functioning whole.

2.2 Entity Relationship Diagram

According to Bagui and Earp (2012), an ERD can be seen as a semantic tool for data modelling, and it has different components, namely; entities, attributes and relationships. An ERD is also a diagram that visualize data in the form of entities representing the tables of a database, the fields are represented by the attributes of an ERD, and it also includes relationships to reflect the relation between tables (Bentley & Whitten, 2007). Through primary keys and foreign attribute keys in the entities the entities are connected to one another, creating relationships between tables. Fig. 1 shows an example ERD.

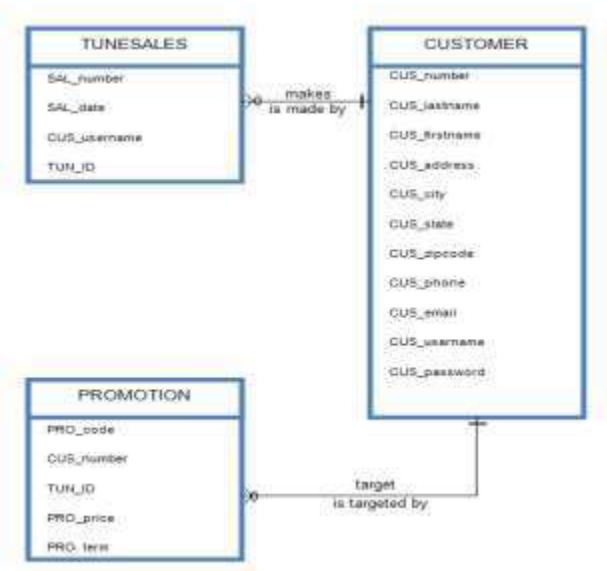

Figure 1: Example ERD adopted from Satzinger, et al.

2.3 Structured Query Language

The language that the third year learners use in their database practical applications, namely Structured Query Language (SQL), runs on the MySQL Workbench environment. MySQL Workbench is free and open-source software.

According to Coronel, Morris, Rob, and Crockett (2013), an SQL engine allows tables to store data and SQL queries to be created to view and retrieve data within a database. This language has many elements such as; statements, queries, predicates, commands and expressions. When interpreting SQL, a conclusion can be made as to what manipulations can be done during the event of query execution (Chapple, 2015). A statement in SQL combines the use of commands, expressions,

predicates and queries. When queries are executed, a record of data returns – depending on the command(s) used within the query. The actions specified by the user and used in the queries are called SELECT statements, which are usually followed by a WHERE clause to identify only certain data (Chapple, 2015). Below, in Table 1, an example query with its results is shown.

Table 1: Example SELECT statement from Taylor (2015)

SELECT * FROM Customer Note: only the first three rows are displayed				
Result				
custID	custName	gender	billAdd	Phone
00001	Dorothy Gale	F	4 Woodmead	016 972 1010
00002	The Wizard of Oz	M	9 Van Eck	016 982 6589
00003	The Scarecrow	M	15 Karee	016 932 1023
…	…	…	…	…

3 SOFTWARE UTILIZED

In this study the artefact was created using Microsoft Visual Studio 2012 integrated development environment, which host numerous programming languages such as Visual Basic .NET, C# and C++. Bell and Parr (2009) indicate that the C# language is used for general development purposes and was the selected programming language for the artefact.

4 METHODOLOGY

For some time Design Science Research (DSR) has been used in the engineering and Information Systems (IS) disciplines (Gregor & Hevner, 2013). Defined within the IS discipline, DSR is the construction of a wide range of socio-technical artefacts such as decision support systems, tools, governance strategies, methods of IS evaluations and IS change interventions (Gregor & Hevner, 2013). DSR also analyses the performance of a designed artefact in order to understand and improve the artefact (Hevner & Chatterjee, 2010; Vaishnavi & Kuechler, 2004). DSR is primarily the creation and evaluation of an artefact used to acquire the solution to the identified problem through understanding thereof. The evaluation of artefacts could be subject to quantitative and/or empirical and qualitative methods (Hevner, March, Park, & Ram, 2004).

The research question to be addressed in this paper is:
"Does the visualization artefact envisaged in support of the learning of databases have value?"

The DSR approach used in this research was suggested by Vaishnavi and Kuechler. Their DSR approach includes; first – becoming aware of the problem and proposing a solution, second – coming up with the idea for a proposed solution which includes a tentative design, third – developing the artefact based on the tentative design as a solution to the problem, fourth – assessing the artefact by stating the artefact's performance measures, and fifth – concluding with results which allows for deciding the artefact's success or determining any further development cycles and research. This approach is discussed in more detail below.

4.1 Problem awareness

The compilation of an ERD is difficult, even more so when a complex scenario is presented. When required to write queries based on an ERD is also difficult at first, especially while the knowledge that there is a relationship or link between these two actions has not dawned on the learner. The problem awareness come from two experience; in the first instance, being a learner in the databases subject modules afforded me a front row seat to the experience of learning about practical database set up and queries, in the second instance being a student assistant allowed me to take a step back and see how many of the learners struggle with the same problems I did. The idea of the creation of an artefact came to mind, and specifically one that allows visualization; something that would allow

learners to instantly see the link between an ERD and its queries and also the manipulation of both to visually reflect actions in the other. At this point it is important to note that the artefact attempts interactive visualization – in the sense that something that is already represented visually (an ERD), is shown interactively for the purpose of supporting understanding.

4.2 Suggested solution

It is well known that learners learn in different ways. From the perspective of Allen, Scheve, and Nieter (2011) a learning style can be seen in a manner of thinking, thoughts being processed to be able to understand difficult concepts. Among other, the most common understanding suggests four leaning styles, namely visual, aural, read/write, and kinaesthetic – making up an acronym, VARK (Fleming & Baume, 2006). At the outset the visualization artefact was imagined to be a learning aid to assist in learners in a number of learning styles, mainly visually – it reflects abstract concepts and links in a visual way, reading and writing – it allows learners to manipulate (write) queries and read the queries the artefact produces, and it allows learners do manipulate ERDs and queries to see the effect the one has on the other. Only the auditory learning style is not addressed in the design of the artefact.

4.3 Artefact development

The visualization artefact is called DBV4L, pronounced Deebee-val. The name reflects the fact that it assists in the leaning of databases (DB), as well as the fact that it is a visual (V) aid designed for (4) learning (L). To facilitate the meaning of its name, two concepts that learners already come across in their second year, namely ERDs and SQL, were used to guide and direct the learners' learning. These two concepts were used throughout DBV4L to meet the tentative design objectives of the research. With the SQL queries and ERDs in mind, an interface was designed to allow learners to work in two directions; from the SQL-statement towards the ERD, as well as and from the ERD towards the SQL-statement. A single screen allows learners to see both windows – with the SQL-statement built on the left half of the screen and the ERD built on the right half of the screen. To start a learning opportunity, a learner has to build the ERD selected for learning utilizing an easy drag-and-drop facility. Pre-requisites for proper functioning regarding the ERD, includes naming the entities, stating each entity's attributes and stating the relationship between entities. This basic interface is shown in Figure 2. Part of the ERD example shown in Figure 1 is recreated in DBV4L to show how it works. The active sections are indicated on the interactive screen; the status line ①, build a query button ②, query window ③, create an entity button ④, graph window ⑤, and create a relationship between entities button ⑥.

From this point it is possible to build an SQL query – DBV4L will then highlight the entities, attributes and relationships involved. While the query is written, highlighting happens interactively. This is depicted in Figure 3. Because of limited space in the figure, the left-hand side of the screen – where the SQL statement is built – is displayed on top of the right-hand side of the screen – where the ERD is built.

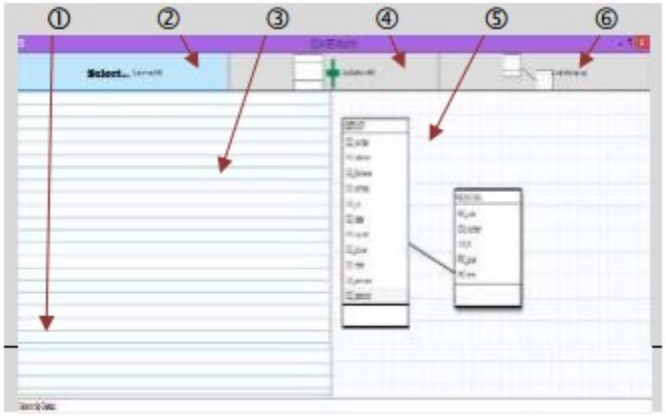

Figure 2: Example of the DBV4L interface

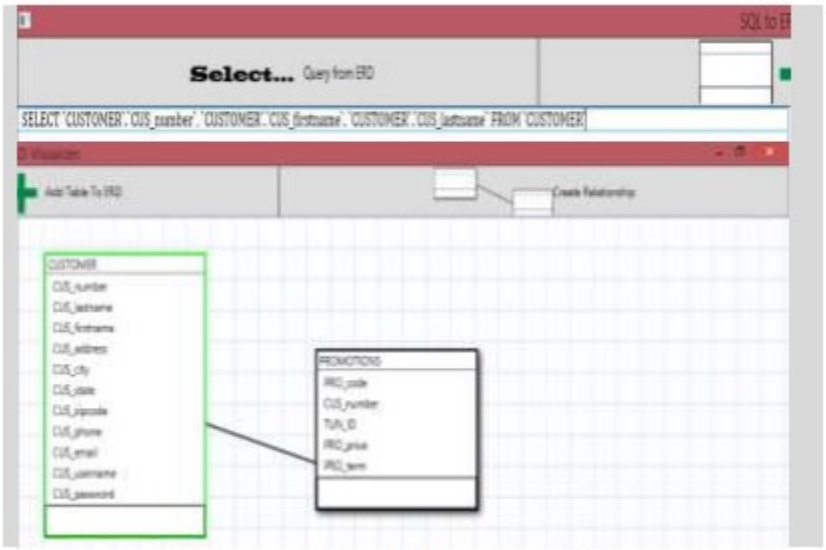

Figure 3: From SQL query to ERD

When a database learner wants to learn how an SQL query is formed, the learner can only highlight the entities, attributes and relationships on the existing ERD and the artefact will build a basic SQL-query. This is happening interactively with the highlighting actions. The representation is shown in Figure 4. In this depiction the order of the display is switched to highlight the order of the implementation.

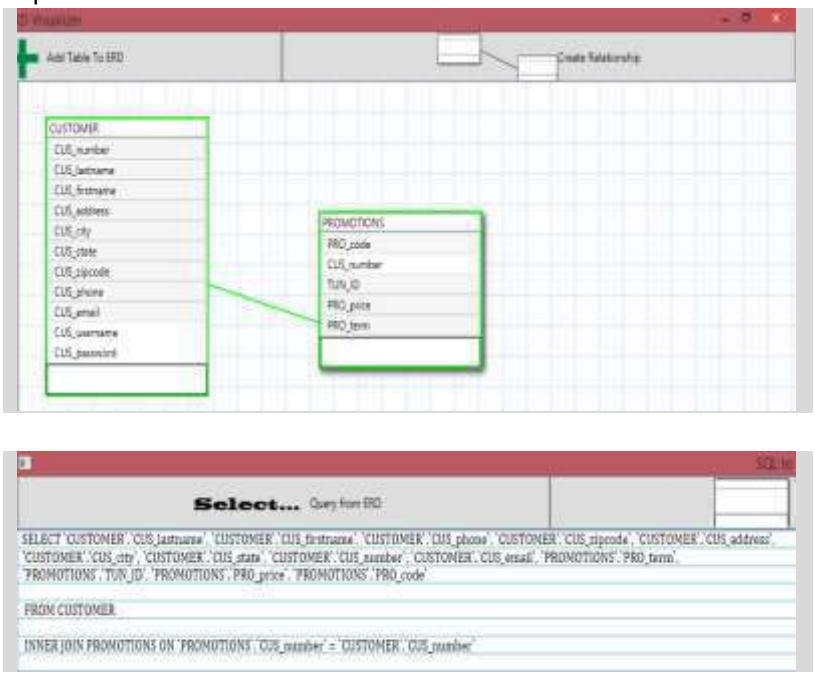

Figure 4: From ERD to SQL query

4.4 Artefact evaluation

It is not guaranteed that the first attempt at a solution will be successful; this can lead to the re-modelling of the research, and the possible re-development of the artefact. With the possibility of the artefact being not feasible, the research must be repeated from the point where the problem was identified and an understanding was formed to enable its analysis (Dresch, Lacerda, & Antunes, 2015). If the artefact has met the requirements of the solution, a conclusion may be drawn and the outcome of the artefact would be labelled as favourable. The results that were obtained from the research will help the researcher to obtain a better understanding in the problem area of the research. For this reason the assessment of the artefact is crucial.

The visualization artefact that was developed had been evaluated and the evaluation was based on the requirements of the artefact being met and it being able to solve the problem. Data was

collected from an interview with an expert – the lecturer. The selection of this person is threefold; the participant is the person teaching the databases subject modules, upon implementation of the artefact, this person is seen as the gatekeeper who needs to be convinced of its value, and this lecturer also has extensive experience in the implementation of databases.

The interview was transcribed and analysed. It revealed that there are positive features to the visualization artefact, including:

1) simplicity of use of the working window – the main page where the SQL queries are being built,

2) ease of use of the drag and drop feature utilized in the graph window – where the ERD is being compiled, and

3) a layout which is neat and clear with no clutter – this may distract or confuse the learner.

According to the participant the visualization artefact is developed well enough for learners to use and learn from it. Features needing more attention include:

1) although a status line indicates what needs to be done next – tooltips which explains the meaning of a feature when the learner hovers a cursor over it will enhance its use,

2) some guidance on how to create the ERD in the graph window – possibly in the form of a help facility, and

3) some guidance regarding where to create the SQL query – this may also be included in a help facility.

As indicated earlier, a number of factors need to be taken into consideration when designing and developing a visualization artefact. In Table 2 an attempt is made to show that all the factors have been taken into account in the development of DBV4L.

Table 2: Factors important in the development of a visualization artefact applied to DBV4L

Factor	Implementation	Source
Flow of events.	Immediate reflection of interaction between ERD and SQL queries	Agosti et al. (2013)
Interaction between a user and the artefact. Presentation of the different components.	The learner uses a drag-and-drop interface, clicking (of elements on the ERD) and text to be typed. The interface is designed in such a way that leaners can easily figure out what is expected from them, although they need to take the first step. The representation of the SQL statement that the leaner built, or DBV4L builds reflect the look and feel of a live SQL statement. The ERD reflected on the screen looks like an ERD prescribed by conventions.	Liang (2010) Ursyn (2014)
Deconstruction of the whole into its components, and then re-constructing it by showing how	When SQL queries are built in DBV4L, the different entities are highlighted to see the smaller parts becoming a whole through the means of the SQL query.	Phillips et al. (2010)

4.5 Artefact conclusion

Since the developed artefact is a visualization aid, the researcher would like to see one important visual implementation, namely that of showing a one-to-many (1:N or 1:*) relationship between two entities. Although this is implemented correctly and working well, its success is based on the link between the primary key and foreign key, which is not visually represented. Otherwise, the main goal of the research, namely to develop a visualization artefact to support database learners in their learning and understanding of the practical implementation of databases, was met.

5 CONCLUSION

The primary goal of this research was to create a visualization artefact to assist learners doing databases subject modules in their learning. The suggested artefact may enhance learners' understanding of the analysis, design, development, and support of databases through visualization. The main focus was on facilitating the understanding of the SQL data manipulation language, but it is anticipated that the understanding of the ERD modelling tool will also be facilitated, especially when learners are not comfortable to use it well. The creation of the DBV4L artefact realized the primary goal. Therefore, from the feedback discussed in the preceding paragraphs, the research question, "*does the visualization artefact envisaged in support of the learning of databases have value*?", may be considered to be answered since most participants answered in the affirmative.

Further discovery and development of DBV4L is possible. A starting point would be to implement the suggestions made by the participant interviewed during the evaluation of DBV4L. In addition to the participants' suggestions, the visual presentation of all possible relationships, including one-to-one (1:1) and one-to-many (1:N or 1:*) – with their cardinalities, would contribute to the usefulness of the artefact. The creation of tables with the possibility to add data and then manipulate the data through the use of DBV4L will add much value when using it as a learning aid.

Further research may include the implementation of DBV4L in the third year database subject modules – where its creation was originally inspired. This will contribute much towards its development, especially when making use of one or more action research (AR) cycles to gather data and to learn from learners using the artefact. Other AR implementations that may be helpful towards the improvement of DBV4L include:

- learners enrolled for second year Systems Analysis & Design subject modules where ERDs and SQL queries are introduced, as well as
- learners enrolled for fourth year honours course which includes two database-based modules, namely Database Systems, and Data Warehouses – utilizing the concepts covered by the artefact and possibly making its inclusion and use valuable.

A combination of the suggested groups may inform the research with regard to where the artefact's value or difficulty is most prevalent.

In one of the suggested AR studies, one or more of the following approaches may be followed:

- Groups of participants may be included in the study, namely industry experts, as well as one or more of the suggested groups of learners. All groups would be required to use the artefact after which they may be interviewed to allow analysis of the findings and eventual understanding of their experience.
- Learning styles, such as VARK that was mentioned earlier, may be used to categorize learners included in the study, and then determine how each group responds to the system.
- A metric may be designed to measure the improvement in understanding the mentioned concepts through the utilization of the artefact. Such an intervention may establish the value of the artefact.
- The difficulty of the use of the artefact may be measured.

ACKNOWLEDGMENT

Edwin Ries and Bernadine Meintjes, who collaborated in the development of DBV4L, is acknowledged for their contribution.

REFERENCES

Agosti, M., Foner, P., Muller, H., & Santucci, G. (2013). Information Retrieval Meets Information Visualisation. London: Springer-Verlag.

Allen, K., Scheve, J., & Nieter, V. (2011). Understanding Learning Styles. California: Shell Educational Publishing, Inc.

Bagui, S., & Earp, R. (2012). Datbase Design Using Entity-Relationship Diagrams. United States of America: Taylor & Francis Group.

Bell, D., & Parr, M. (2009). C# for Students. England: Pearson Education Limited.

Bentley, L. T. M., & Whitten, J. L. (2007). Systems Analysis and Design for Global Enterprise. New York: McGraw-Hill Companies.
Chapple, M. (2015). SQL Lurks Everywhere. Retrieved from http://databases.about.com/od/sql/a/What-Is-Sql.htm.
Coronel, C., Morris, S., Rob, P., & Crockett, K. (2013). Database Principles.
Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). Design Science Research. Switzerland: Springer International Publishing.
Fishman, B. J., & O'Conner-Divelbiss, S. F. (2000). Facing the Challenges nof Complex Real-World Settings. Fleming, N., & Baume, D. (2006). Learning Styles Again: VARKing up the right tree! Educational Developments, 7, 4.
Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS Quaterly, 37(2), 337-355.
Hevner, A. R., & Chatterjee, S. (2010). Design Research in Information Systems (Vol. 22). New York Dordrecht Heidelberg London: Springer.
Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. MIS Quaterly, 28(1), 75-105.
Liang, H. (2010). An Exploratory Study of Interactivity in Visualisation Tools: 'Flow' of Interaction.
Phillips, L. M., Stephen, N., & Norris, J. S. (2010). Models and Modelling in Science Education: Visualisation in Mathematics, Reading and Science Education. London: Springer Media.
Taylor, A. G. (2015). SQL WHERE Clause Predicates.Retrieved from http://www.dummies.com/how- to/content/sql-where-clause-predicates.html
Ursyn, A. (2014). Perceptions of Knowledge Visualisation: Explaining Concepts through Mathematical Images. United States of America: IGI Global.
Vaishnavi, V., & Kuechler, K. (2004, 16 August 2009). Design Research in Information Systems. Retrieved from http://desrist.org/design-research-in-information-systems/