International Journal of Research in Advanced Engineering and Technology
Online ISSN: 2455-0876
www.engineersjournals.com
Volume 1; Issue 3; December 2015; Page No. 46-48

Quality metrics: A literature review

Naveen Dahiya

Department of CSE, MSIT, New Delhi, India

Abstract
The paper presents a brief overview of the work carried out by researchers in the domain of quality metrics. The quality metrics play significant role in quality evaluation of data warehouse models at conceptual and logical level. The quality metrics reflect the size and structural properties of data warehouse models.

Keywords: Quality metrics, conceptual, logical, data warehouse.

1. Introduction

Data warehouse store large amounts of information about organizations and play an important role in handling of fierce business competition by analysis of past trends towards prediction of future trends. To enable fast extraction of useful information the quality of data warehouses has to be maintained. One of the major factors that help in quality evaluation of data warehouses is quality metrics. The focus of the paper is on the study of various quality metrics proposed by researchers and discussing their role in quality evaluation of data warehouse models.

2. Literature Review

The design of a data warehouse system is an incremental process starting with conceptual phase, then logical phase and finally physical phase. The conceptual design phase is the most important phase of the data warehouse design process as it lays the foundation for effective software development. The quality of a data warehouse at conceptual and logical level can be evaluated using quality metrics that are based on size and structural properties of models. Several quality metrics have been proposed by researchers to evaluate the quality of conceptual and logical models of data warehouse development. The related literature is as follows:

Serrano *et al*., 2007 [1] has proposed quality metrics for quality evaluation of conceptual models. The proposed metrics are theoretically and empirically validated to prove their utility. The metrics proposed by him for a schema S are as follows:

- NDC(S) Number of dimension classes.
- NBC(S) Number of base classes.
- NC(S) Total number of classes, NC(S) = NDC(S) + NBC(S) + 1
- RBC(S) Ratio of base classes. Number of base classes per dimension class.
- NAFC(S) Number of FA attributes of the fact class.
- NADC(S) Number of D and DA attributes of the dimension classes.
- NABC(S) Number of D and DA attributes of the base classes.
- NA(S) Total number of FA, D and DA attributes, NA(S) = NAFC(S) + NADC(S) + NABC(S)
- NH(S) Number of hierarchy relationships DHP(S) Maximum depth of the hierarchy relationships.
- RSA(S) Ratio of attributes. Number of attributes FA divided by the number of D and DA attributes.

Empirical validation was applied to prove that several of the metrics seemed to be practical indicators of conceptual model understandability.

Serrano *et al*., 2008 [2] proposed structural metrics for quality evaluation of logical models and carried out an empirical study to investigate their significance in quality evaluation of logical models. The proposed metrics are as follows:

- NFT (Sc). Number of fact tables in the schema.
- NDT (Sc). Number of dimension tables in the schema.
- NFK (Sc). Number of foreign keys in all the fact tables of the schema.
- NMFT (Sc). Number of facts in the fact tables.

Genero *et al*., 2007 [3] proposed 3 size metrics and 8 structural metrics for conceptual models as follows:

Size metrics

- Number of Classes (NC) The total number of classes in a class diagram.
- Number of Attributes (NA) The number of attributes defined across all classes in a class diagram (not including inherited attributes or attributes defined within methods).
- Number of Methods (NM) The total number of methods defined across all classes in a class diagram, not including inherited methods.

Structural metrics

- Number of Associations (NAssoc) The total number of association relationships in a class diagram
- Number of Aggregations (NAgg) The total number of aggregation relationships (each “wholepart” pair in an aggregation relationship).
- Number of Dependencies (NDep) The total number of dependency relationships.
- Number of Generalizations (NGen) The total number of generalization relationships (each “parent-child” pair in a generalization relationship).

- Number of Generalization Hierarchies (NGenH) The total number of generalization hierarchies, i.e. it counts the total number of structures with generalization relationships.
- Number on Generalization Hierarchies (NAggH) The total number of aggregation hierarchies, i.e. it counts the total numbers of “whole-part” structures within a class diagram.
- Maximum DIT (MaxDIT). The maximum DIT value obtained for each class of the class diagram. The DIT value for a class within a generalization hierarchy is the longest path from the class to the root of the hierarchy
- Maximum HAgg (MaxHAgg) The maximum HAgg value obtained for each class of the class diagram. The HAgg value for a class within an aggregation hierarchy is the longest path from the class to the leaves.

Basili *et al*., 1999 [4] discussed the importance of experimentation in empirical validation. The main focus was on defining goal and hypothesis, experimentation and validity threats to experiments.
Calero *et al*., 2001 [5] proposed various metrics for different configurations of data warehouse schemas.

- Table metrics: NA, NFK
- Star metrics: NDT, NT, NADT, NAFT, NA, NFK, RSA, RFK
- Schema metrics: NFT, NDT, NSDT, NT, NAFT, NADT, NASDT, NA, NFK, RSDT, RT, RFK, RSDTA

Also these metrics were theoretically validated to prove their relevance.
Shull *et al*., 2008 [6] focused on the role of replications in empirical study. Replications were categorized in two types:

- Exact replication
- Conceptual replication

Exact replication was further classified as dependent and independent replication. Goals, benefits, limitations of each were discussed with due emphasis on documentation.
Moody, 2005 [7] discussed various theoretical and practical issues in evaluating the quality of conceptual models with special emphasis on experimental techniques.
Gosain *et al*., 2011 [8] conducted a replica study to explore a correlation between understandability and metrics proposed by Serrano *et al*., 2008. The results show that NFT, NDT, NFK have significant role towards predicting understandability of logical schemas, while NMFT was not found to be correlated to understandability. Also the combined effect of different combinations of metrics using univariate and multivariate regression was carried out.
Lucia *et al*., 2010 [9] conducted three sets of controlled experiments aimed at analyzing whether UML class diagrams are more comprehensible than ER diagrams during data models maintenance. The results indicated that UML class diagram subjects achieved better comprehension levels.
Hofman, 2011 [10] conducted an empirical validation to analyze the ‘history effect’ in software quality evaluation process. A simplified method was proposed to manipulate observed quality level for a product, thereby making it possible to conduct research. The results showed significant negative influence of negative experience of users on final opinion about software quality regardless of its actual level.
kpodjedo *et al*., 2011 [11] performed an investigation to find the usefulness of elementary design evolution metrics to identify defective classes. It was shown that design evolution metrics make significantly better predictions of defect density than other metrics and thus help in reducing testing effort by focusing test activity on reduced volume of code.
Haigh, 2010 [12] conducted an empirical study involving survey of more than 300 current and just graduated students asking them to rate the importance of 13 quality attributes related to software. The results showed differences in some but agreement in many areas.

3. Research Implications

The paper presents various quality metrics proposed by several researchers and their relative importance in quality evaluation of data warehouse models. The study of existing literature gives a clear understanding of quality metrics and their applications. The basic knowledge about quality metrics can be used by researchers to propose new metrics for quality evaluation of data warehouse models and prove their utility by conducting theoretical as well as empirical validation. As more relevant metrics are discovered, more accurate will be the quality evaluation of data warehouse models.

4. Conclusion

In the paper, several quality metrics and their role in quality evaluation of data warehouse models is discussed. The current literature cannot be said to be exhaustive due to several limitations of accessibility and time constraints. The literature can be further enhanced by discovering new relevant quality metrics and predicting their role in quality evaluation of data warehouse models.

5. References

1. Serrano M, Trujillo J, Calero C, Piattini M. Metrics for data warehouse conceptual models understandability, Information and Software Technology 2007; 49(8):851-870.
2. Serrano MA, Calero C, Sahraoui HA, Piattini M. Empirical studies to assess the understandability of data warehouse schemas using structural metrics. Software Quality Journal.2008; 16(1):79-106.
3. Genero M, Poels G, Piattini M. Defining and validating metrics for assessing the understandability of entity–relationship diagrams, Data & Knowledge Engineering, 2008; 64(3):534-557.
4. Basili V, Briand L, Melo W. A validation of object-oriented design metrics as quality indicators, IEEE transactions software engineering 1999; 22(10):751-61.
5. Calero C, Piattini M, Pascual C, Serrano M. Towards Data warehouse Quality Metrics, International Workshop on Design and Management of Data Warehouses (DMDW’01), 2001.
6. Shull F, Jeffrey Carver C, Sira Vegas, Natalia Juristo. The Role of Replications in Empirical Software Engineering. Journal of Empirical Software Engineering Springer. 2008; 13:211-218.
7. Moody D. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future

directions, Data & Knowledge Engineering 2005; 55(3):243-276.
8. Gosain A, Sabharwal S, Nagpal S. Assessment of quality of data warehouse multidimensional model. Int. J. of Information Quality. 2011; 2(4):344-358.
9. Lucia A, Carmine Gravino, Oliveto R, Tortora G. An experimental comparison of ER and UML class diagrams for data modeling. Journal of empirical software engg. Springer. 2010; 15:455-492.
10. Hofman R. Behavioral economies in software quality engineering. Journal of empirical software engineering. 2011; 16(2):278-293.
11. Kpodjedo S, Ricca F, Galinier F, Gueheneuc Y, Antoniol G. Design evolution metrics for defect prediction in object oriented systems, Empirical Software Engineering 2011; 16(1):141-175.
12. Haigh M. Software Quality, non-functional software requirements and IT-business alignment. Software Quality Journal. 2010; 18(3):361-385.