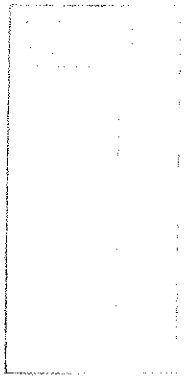
Modeling and Analysis of Metrics Databases

by

Raymond A. Paul

A Dissertation Presented To
The Faculty Of University Of Tsukuba
In Candidacy For The Degree Of
Doctor Of Engineering

January 1999



This dissertation is dedicated to the memory

of my mother, Bertha Paul (1917-1986)

ABSTRACT

In this thesis we investigate important research issues related to quantitative project management in software engineering and propose various approaches to tackle these challenges. Metric databases are among the most useful resources available to project a mangers to aid in their decision making. We identify a core set of software metrics to be collected for a wide range of projects. These metrics are broadly classified as management, reliability, and quality metrics. Proper analysis of these software metrics and correct interpretation of the analysis results can provide comprehensive information to aid in management decision-making, with the consequent reduction of risks associated with software development projects. For this purpose we propose a framework for integrating various statistical and analytical techniques with software metrics data. The proposed framework can provide a powerful environment for risk monitoring and assessing the impacts of corrective actions and decisions made by the management team.

One of the key contributions of this research is development of two models that can be used to describe the productivity of a software team as a function of effort. A comprehensive analysis of these models based on real data provides a strong validation of these models. The collection of metrics data for past projects and their integration into a metrics database enables us to accumulate past experiences on previous projects and reuse them for analysis of current and planned projects. Management can then pose queries on the metrics database using the appropriate database query language or request

for further analysis of raw metrics. The underlying data model for the DBMS has a direct impact on the ease and efficiency with which such queries are posed, and we therefore examine the effectiveness of different data models with respect to various classes of database queries. With respect to the DBMS data model, we compare various data models and conclude that while the relational model proved adequate for many simple and straightforward queries, alternative models such as object-oriented and graph data models perform better when more complex queries are posed; particularly with queries involving temporal dimension and recursion. Moreover, the metrics database may undergo structural reorganizations for efficiency or other reasons as metrics data for new projects are added to the database and as different queries are identified. The relational model is relatively inflexible to structural modification as it requires the queries to be rephrased; the object-oriented and graph data model, however, does not suffer from such restrictions. The choice of the underlying data model for a metrics database therefore depends very much on the queries the management expects to pose, as well as the type of analysis to be performed.

In order to support broad range of semantic queries we propose a formal framework based on a recursive graph formalism that uses a a Petri Net based model. The proposed framework can allow management of software data schema evolution as well as can assist user to construct arbitrary object-oriented views of risk/quality associated with a software project.

ACKNOWLEDGMENTS

I would like to express my heartfelt thanks to my adviser Professor Nobuo Ohbo for providing encouragement, numerous suggestions and guidance on this research. I am thankful to all the Professors on my dissertation committee, for providing me insight into several research ideas. I am also grateful to Professor T. Kunii for his comments on the relational and graph data models and providing me a clear understanding of the importance of the underlying data model with respect to metrics databases

My interest in metrics databases arise from my professional responsibilities at the US Department of Defense, and my numerous interactions concerning metrics with software engineers and program managers of major multinational engineering corporations. Thanks, too, go to my colleagues for their insight as to the importance of metrics databases in project management, and their contributions on new queries and analysis techniques for metrics databases. My interaction with them in these areas has been most helpful in my research.

Professor C. V. Ramamoorthy at the University of California, Berkeley, Professor Y. Shinagawa, Prof. A. Goel at Syracuse University, and Professor A. Ghafoor at Purdue University, have also been most helpful in providing me with additional insight as to related work on software project management and metrics databases, as well as performance comparison techniques for various DBMS data models. I would like to take this opportunity to thank them for their valuable time and effort.

Last but not least, I am most grateful to my wife Alice and my daughter Ann for the concern, understanding, and moral support they have provided me throughout my work.

TABLE OF CONTENTS

Appendix B: NSDIR and NEC Data Analysis

LIST OF FIGURES

LIST OF TABLES **Page**