SHARING AND PRESERVING CODING BEST PRACTICES THROUGH PROGRAMMER DATA ANALYTICS

Samiththa Bashani

(168210M)

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2018

SHARING AND PRESERVING CODING BEST PRACTICES THROUGH PROGRAMMER DATA ANALYTICS

Jasing Pathiranage Samiththa Bashani

(168210M)

Thesis submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science and Engineering

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

June 2018

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books)

Signature: ....   Date: ..….......................

J.P. Samiththa Bashani

The above candidate has carried out research for the Masters thesis under my supervision.

Name of the supervisor: Dr. Indika Perera

Signature of the supervisor: ....   Date: ....

Abstract

Sharing and preserving coding best practices among the developers are becoming an important objective of software development life cycle. Because violations on coding best practices may lead to catastrophic events which are costly and time consuming. There has been numerous researches done in order to mitigate the issues related to bad coding practices. One of the most challenging tasks towards mitigating this is to identify the skill level of the developers, coding patterns and likelihood for bad coding practices. The widely used methods for this are conducting one on one interview with the developers and review developers work.

This particular research tried to contribute to the field of software architecture by analyzing the feasibility of using machine data to identify the developer coding patterns and related data and provide a mechanism to enhance the skills of a developer. By doing that it makes sure an organization can share and preserve the coding best practices within an organization.

This research focused on developing a parsing mechanism to collect those data from various file formats and types. For this research scope it focused on the static code analysis tool called FindBugs and log data. A successful parser of logs formatted in XML has been developed. A central data storage architecture has been developed in order to capture data from various sources which are different from each other.

Collected data analyzed to generate information about the developers' pattern in doing mistakes and coding styles. To prove that analyzing programmer data for a significant period can predict their abilities and weaknesses an evaluation has been carried out. The evaluation compare data from developer spot interviews with developers' log analyzed data. With those comparisons it identified log data results can match the interview results in an 80% success rate.

ACKNOWLEDGEMENT

I would like to express my profound gratitude and deep regards to my supervisor Dr. Indika Perera for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout this work. Also I am grateful for the support and advice given by Dr. Malaka Walpola, by encouraging this research.

Finally I would like to thank the academic and nonacademic staff of Department of Computer Science and Engineering and collogue of MSC'16 for the support and encouragement given.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
SCA	Source Code Analyzer
SIG	Software Improvement Group
SDLC	Software Development Life Cycle
SDK	Software Development Kit
IDE	Integrated Development Environment