المنهل
ALMANHAL

نبع المعرفه
THE SOURCE OF
KNOWLEDGE

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law
https://platform.almanhal.com/Details/Thesis/81815

Actual dissertation cover is not available   غلاف الرسالة غير متوفر

Powered by

REVERSE SOFTWARE ENGINEERING AND REENGINEERING TO DETECT PLAGIARISM IN JAVA PROGRAMS

By

Khair Eddin Muawiyah Sabri

Supervisor

Dr. Jubair J. Al-Ja'afer, Prof

This Thesis was submitted in Partial Fulfillment of the Requirements for the Master's Degree of Science in Computer Science

Faculty of Graduate Studies

The University of Jordan

December, 2003

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

This Thesis (Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs) was successfully defended and approved on December 4, 2003

Examination Committee	Signature
Dr. Jubair J. Al-Ja'afer, Supervisor Prof. of Software Engineering	-------------------------
Dr. Eyas Abdel-Rahim El-Qawasmeh, Member Assistant Prof. of Object-Oriented, Information retrieval on the web, and Error-correction.	-------------------------
Dr. Fawaz Ahmed Masoud, Member Associate Prof. of Software Engineering	-------------------------
Dr. Amjad Ahmed Hudaib, Member Assistant Prof. of Software Engineering	-------------------------

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

Dedication

To my lovely parents for their valuable support and motivation,

without which this work might not have been achieved.

Also, to my brother Zahir and sisters Suha and Dana

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

ACKNOWLEDGEMENT

It is my pleasure to extend high appreciation and recognition to my advisor Dr. Jubair J. Al-Ja'afer for his valuable guidance, fellow-up and support.

I sincerely thank the examination committee members: Dr. Eyas El-Qawasmeh, Dr. Fawaz Masoud and Dr. Amjad Hudaib for their significant suggestions and comments.

I highly appreciate and acknowledge the cooperation of all members of the King Abdullah II School for Information Technology.

Finally, a special thanks to the University of Jordan for granted me a scholarship to attain the master degree in computer science.

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

List of Contents

Subject	Page
Committee Decision	ii
Dedication	iii
Acknowledgement	iv
List of Contents	v
List of Tables	viii
List of Figures	ix
Abstract	x
Introduction	**1**
1 Reverse Engineering and Reengineering	1
2 Information Content of a Program (Entropy)	2
3 Detecting Plagiarism	3
4 Automated Program Evaluation	4
5 Objective of the study	5
Literature Review	**6**
1 Reverse Engineering and Reengineering	6
2 Plagiarism	7
3 Software Measurement	10
3.1 The IEEE Definitions	12
3.2 Traditional Software Measurements	12
3.3 Object Oriented Measurements	15
3.4 STYLE Metrics	19
4 Automated Program Evaluation	20
Proposed work	**23**
1 Detecting Plagiarism	23
1.1 Program Structure	23
1.2 Performing the Analysis	26
1.2.1 Inspecting the General Characteristics	26

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

Subject	Page
1.2.1 Measuring the Entropy	27
1.3 Algorithm	29
2 Assessment of Design Quality	32
2.1 CK Metrics	32
2.2 LK Metrics	34
2.3 The MOOD Metrics	35
2.5 STYLE Metrics	41
2.5 Algorithms.	46
2.5.1 CK & LK Metrics	46
2.5.2 The MOOD Metrics	50
2.5.3 The STYLE Metrics	52
2.5.4 The Algorithm to Compute Final Grade	56
System Design, Implementation, Application, and Result Analysis	**59**
1 System Design and Implementation	59
1.1 System Design	59
1.2 System Implementation	62
2 System Application	65
2.1 Detecting Plagiarism	65
2.2 Programs Evaluation	65
3 Results and Analysis	66
3.1 Detecting Plagiarism	66
3.2 Programs Evaluation	71
Conclusions and Future work	**81**
1 Conclusions	81
1.1 Detecting Plagiarism	81
1.2 Evaluation	82
2 Suggested Future work	83
References	**84**

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

Subject Page

Appendix A Detsting Plagiarism Using JK System.. 91

Appendix B Evaluation of Different Designs Using JK System...................... 98

Appendix C Evaluation of Simple Java Programs Using JK System........... 129

Abstract in Arabic 136

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

List of Tables

Table Number	Table Title	Page
1	The importance of CK & LK factors	49
2	The range of MOOD factors	50
3	The importance of MOOD factors	52
4	The importance of STYLE factors	56
5	The similarity based on Entropy case 1	67
6	The similarity based on Entropy case 2	68
7	The similarity based on Entropy case 3	69

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

List of Figures

Figure Number	Figure Title	Page
1	Examples of equivalence classes	3
2	AUTOMARK approach to mark FORTRAN programs	21
3	JK System main functions	23
4	The process of generating Terminating Binary Sequence	24
5	Flow chart of the detecting plagiarism algorithm	31
6	The JK metric suit	46
7	Block diagram of the JK system	61
8	The distribution of student marks using JK	76

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

REVERSE SOFTWARE ENGINEERING AND REENGINEERING TO DETECT PLAGIARISM IN JAVA PROGRAMS

By

Khair Eddin Muawiyah Sabri

Supervisor

Dr. Jubair J. Al-Ja'afer, Prof

ABSTRACT

In this research, a system, referred to as Jubair-Khaireddin (JK), has been developed to assess the degree of similarity between two programs even though they look superficially not related. JK system has the capability to detect deliberate attempts of plagiarism. Also, JK system for used in program evaluation.

This research used reverse engineering and reengineering to detect plagiarism. Reverse engineering is used to bring each program back to its initial specification stage. This methodology enables us to extract the structure of the program which is considered an important factor to detect plagiarism. This can be done by constructing the Static Execution Tree (SET) for each program. The SET is then transformed into

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

Terminating Binary Sequence (TBS). The TBSs generated from the tested programs are compared in order to get similar branches. Reengineering technique is applied on these similar branches to compute its Entropy in order to measure the similarity between the branches. The system has been tested on different Java programs with different modifications, and proved its success in detecting almost all cases including those of partially plagiarized programs

Since the quality assessment of a program is a critical process to ensure its effectiveness, JK system is used for program evaluation. This system automatically marks Java programs based on JK metric suite which includes: Chidamber and Kemerer (CK), Lorenze and Kidd (LK), Metrics for Object Oriented Design (MOOD) and STYLE Metrics. The JK system, when used by academics, saves the instructors' time, and provides feedback to both the instructor and students. The system can also be used by programmers to evaluate industrial programs.

The STYLE Metric is an improved version of Redish and Smyth tool called Automark. In JK system, two improvements have been made to Automark. The first is the introduction of new factors to give the tool flexibility in evaluating object-oriented languages such as Java. The second is the automatic generation of a model template for a program under evaluation by the tool, instead of writing a specific model to evaluate each program. JK system has been tested on different programs, and the results proved the effectiveness of the tool in identifying the deficiencies of a program.

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

Introduction

1 Reverse Engineering and Reengineering

A common problem experienced by the software engineering community is the understanding of legacy code. Legacy code is a semiformal term that refers to the programs coded for typical industry projects, which become more difficult to understand as they grow in size and complexity (Pressman, 2000).

Reverse engineering is the process of analyzing a system to identify its components and their relationship. Reverse engineering creates representations of the system in another form or at a higher level of abstraction. With the rapid expansion of software development, the cost of software maintenance rises dramatically. Such cost usually covers 50%—90% (Sommerville, 2001) of the entire software budget. Therefore, reverse engineering tools are important in software maintenance due to the effectiveness of reverse engineering in understanding programs and analyzing the consistency between design and implementation (Sommerville, 2001).

Almost in all cases, reverse engineering is followed by reengineering phase to regenerate the code. Software reengineering is concerned with the reimplementation legacy systems to make them more maintainable. Reengineering may involve redocumenting, reorganizing and restructuring the system. However, the functionality of the system doesn't change, and usually its architecture remains almost the same (Pressman, 2000).

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

2 Information Content of a Program (Entropy)

One of the most fundamental results of the statistical communication theory is the Shannon's information theory (Mohanty, 1979; Shooman, 1983). By ignoring the meaning of a message and focusing on the probability of choosing any symbol out of the message, Shannon was able to establish an Entropy function which measures the statistical information content. By applying the Shannon's concept to a program, its Entropy can be calculated by using the following formula (Shooman, 1983):

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$

Where:

p_i: probability of occurrence of the ith symbol in a program.

n: total number of symbols.

It should be noted that the value of Entropy is always positive. Therefore, a negative sign is applied to the beginning of the formula to change the sign generated from the log function from negative to positive.

The Entropy software metric, discussed by Davis and Leblanc (1988), explains the notion of Entropy at a higher conceptual level by considering the so-called chunks of code. A chunk could be a single statement, a block of code, or a module itself. An important notion is an equivalence class of chunks. The concept of the equivalence class is based on chunks' in-degree and out-degree. Two chunks are considered equivalent if they have the same number of in-degree and out-degree of links with the environment. For example in Figure 1, A and C fall into the same equivalence class; similarly B and E are placed in the same equivalence class (Peters and Pedrycz, 2000).

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815

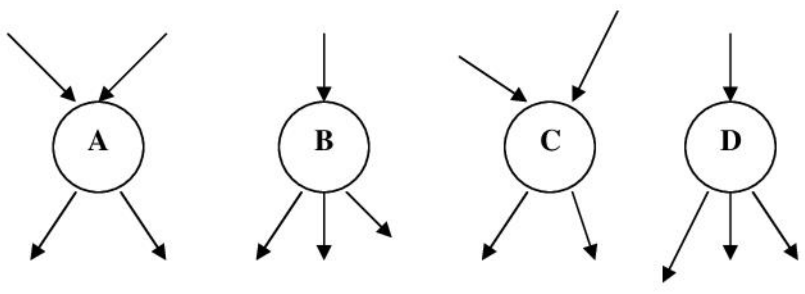

Figure 1: Examples of equivalence classes.

3 Detecting Plagiarism

Students of various disciplines, especially Computer Science, develop numerous numbers of programs worldwide every year as part of their learning process. In most cases, their instructors have the feeling that some of these programs are copied from other programs after some modifications. Copying and claiming others' work partially or completely is called plagiarism.

Traditional search for plagiarism is ineffective and time-consuming. The instructor may discover plagiarism incidentally by observing that a student has forgotten to replace the name of his friend in the program source text, or if two programs show the same failure for a test input. Therefore, a powerful automated search tool that can find similar pairs among a set of programs would be effective and practical.

In this research, the proposed system is intended to detect deliberate attempts of plagiarism, so that it can be used by the instructors to check the copy among students. The JK system also has a wider application, as it offers a means of characterizing programming style and attaching a "Fingerprint" to the program. Another incentive for developing JK system is the protection of intellectual property. Currently software

All Rights Reserved - Library of University of Jordan - Center of Thesis Deposit

Reverse Software Engineering and Reengineering to Detect Plagiarism in Java Programs = الهندسة البرم
Sabri , Khair Eddin Muawiyah
Al Manhal Platform Collections (https://platform.almanhal.com) - 22/11/2024 User: @ Al Manhal Demo Library
Copyright © University of Jordan. All right reserved.
May not be reproduced in any form without permission from the publisher, except fair uses permitted under applicable copyright law.
https://platform.almanhal.com/Details/Thesis/81815