

PLAGIARISM DETECTION SYSTEM FOR JAVA PROGRAMMING  
ASSIGNMENTS BY USING GREEDY STRING-TILLING ALGORITHM

NORULAZMI BIN KASIM

UNIVERSITI UTARA MALAYSIA  
2008





**KOLEJ SASTERA DAN SAINS**  
**(College of Arts and Sciences)**  
**Universiti Utara Malaysia**

**PERAKUAN KERJA KERTAS PROJEK**  
**(*Certificate of Project Paper*)**

Saya, yang bertandatangan, memperakukan bahawa  
(*I, the undersigned, certify that*)

**NORULAZMI KASIM**

calon untuk Ijazah  
(*candidate for the degree of*) **MSc. (ICT)**

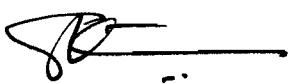
telah mengemukakan kertas projek yang bertajuk  
(*has presented his/her project paper of the following title*)

**PLAGIARISM DETECTION SYSTEM FOR JAVA PROGRAMMING**  
**ASSIGNMENTS BY USING GREADY-STRING-TILING ALGORITHM**

seperti yang tercatat di muka surat tajuk dan kulit kertas projek  
(*as it appears on the title page and front cover of project paper*)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan  
dan meliputi bidang ilmu dengan memuaskan.  
(*that the project paper acceptable in form and content, and that a satisfactory*  
*knowledge of the field is covered by the project paper*).

Nama Penyelia Utama  
(*Name of Main Supervisor*): **MR. MOHD. SHAMRIE BIN SAININ**

Tandatangan  
(*Signature*) : 

Tarikh  
(*Date*) : **22/5/08**

PLAGIARISM DETECTION SYSTEM  
FOR JAVA PROGRAMMING ASSIGNMENTS  
BY USING GREEDY STRING-TILLING ALGORITHM

A thesis submitted to the Graduate School in partial  
Fulfillment of the requirement for the degree  
Master of Science (Information and Communication Technology)  
Universiti Utara Malaysia

By  
Norulazmi bin Kasim

## **PERMISSION TO USE**

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of the Graduate School. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

Dean of Graduate School  
Universiti Utara Malaysia  
06010 UUM Sintok  
Kedah Darul Aman.

## ABSTRACT

Plagiarism is a serious problem faced by universities in the world. University Utara Malaysia, just like any other universities, also have to face this problem. Among Information Technology graduates, Java programming assignments is an essential part of learning programming as it trains the student to solve programming assignments so that they can improve their programming skills that is useful in their professional life after graduation. Unfortunately, some students resolve to copy their assignment from their friends, thus perform the act of plagiarism. For lecturers to detect plagiarism manually is tedious and time consuming, and in most cases almost impossible to be done. This project aims at developing a prototype software tool that can help lecturers detect similarities among student Java programming assignment, thus help lecturers to identify plagiarism offenders among students. The prototype system, known as Java Plagiarism Detection System (JPDS), implements the Greedy-String-Tiling algorithm to detect similarities among tokens in a Java source code files. It is hoped that this prototype will be the stepping stone for development of a more robust plagiarism detection software tool in the future.

## **ACKNOWLEDGEMENT**

Alhamdulillah, most gracious thanks to Allah S.W.T, for His blessing that enables me to work and complete this project.

My first and foremost thanks goes to my supervisor, En. Mohd Shamrie Sainin, for his help, guidance and ideas in making the completion of this project possible. May he and his family is blessed with Allah S.W.T rahmah always.

I would also like to thank all my friends, especially Miss Irene and En. Mohd Sallehuddin for their kind support and help and to all the staff at IT department for their persistent assistance.

## **DEDICATION**

This project is dedicated to my wife, **Salmah binti Katap**, and my new born son, **Muhammad Mifdhal bin Norulazmi**, who was born on the 28<sup>th</sup> May, 2008. You are the inspiration of my life....

## TABLE OF CONTENT

	Page
PERMISSION TO USE	1
ABSTRACT	II
ACKNOWLEDGMENTS	III
DEDICATION	IV
LIST OF TABLE	VII
LIST OF FIGURES	VII
CHAPTER 1: INTRODUCTION	1
1.1    Plagiarism in Programming assignments	2
1.2    Problem statements	3
1.3    Research objectives and significant of study	4
CHAPTER 2: LITERATURE REVIEW	6
2.1    SIM	6
2.2    YAP and YAP3	7
2.3    JPlag	8
2.4    MOSS	8
2.5    Other techniques	9
CHAPTER 3: DETECTION PROCESS	11
3.1    Common practices of programming plagiarism	11
3.2    Tokenizing the files	13
3.3    The comparison process	14
3.4    Calculation of similarity value	17
CHAPTER 4: METHODOLOGY	19
4.1    System scope	21
4.2    Functional requirements of the system	22
4.3    Use case specification	23
4.3.1    Use case diagram	23
4.3.2    Overall flow of events	24
4.4    Analysis and design specification	26
4.4.1    Sequence diagram	26
4.4.2    Class diagram	29

4.5 Implementation	31
CHAPTER 5: RESULTS	
5.1 The students upload files module	32
5.1.1 Changing the upload folder	34
5.1.2 Changes made to the files upon upload	36
5.2 The file comparison module	38
5.3 The test results	44
5.3.1 Testing with mock files	44
5.3.2 Testing with actual student assignment	46
CHAPTER 6: CONCLUSION	
BIBLIOGRAPHY	
APPENDIX A: MOCK FILE FOR SYSTEM TEST	53
APPENDIX B: ACTUAL STUDENT ASSIGNMENT FILES FOR SYSTEM TEST.	56

## LIST OF TABLES

	Page
Table 1: Software used in JPDS development	31
Table 2: Selected comparison output from the mock files test	45
Table 3: Summary of comparison done on eight actual student assignments	47

## LIST OF FIGURES

	Page
Figure 1: Pseudo code for the Greedy-String-Tiling algorithm	15
Figure 2(a): Development phases of the application	19
Figure 2(b): Development cycle for each phase	20
Figure 3: The complete use case diagram for JPDS	24
Figure 4: Sequence diagram for ‘Submit files’ use case	26
Figure 5: Sequence diagram for ‘Open new project’ use case	27
Figure 6: Sequence diagram for ‘Open project’ use case	28
Figure 7(a): Partial class diagram for JPDS	29
Figure 7(b): Partial class diagram for JPDS	30
Figure 8: The directory structure of the Apache Tomcat 6.0 server	33
Figure 9: Web page for uploading files	33
Figure 10: The response page upon file upload	34
Figure 11: The web.xml content	35
Figure 12: Source code for FileUploadHTML file	37
Figure 13: Initial interface of JPDS	39
Figure 14: The ‘Create new JDS project’ dialog box	40
Figure 15: Standard open dialog box for selecting files from local computer	40
Figure 16: The ‘FTP File Downloads’ to select files on FTP server	41
Figure 17: JPDS output example	42
Figure 18: File viewer shows two files that has been compared	43
Figure 19: The output of mock test files comparison	44
Figure 20: The output of comparison made to eight student assignments	47

## CHAPTER 1

### INTRODUCTION

Wikipedia [1] defines plagiarism as “the practice of claiming or implying original authorship of (or incorporating material from) someone else's written or creative work, in whole or in part, into one's own without adequate acknowledgement”. Plagiarism.org [3] further specified the following practices as plagiarism:

- a) turning in someone else's work as their own
- b) copying words or ideas from someone else without giving credit
- c) failing to put a quotation in quotation marks
- d) giving incorrect information about the source of a quotation
- e) changing words but copying the sentence structure of a source without giving credit
- f) copying so many words or ideas from a source that it makes up the majority of their work, whether credit is given or not

Plagiarism is a plagued that is spreading through out academic world and endangering the credibility and quality of graduates as well as the academic institution itself. The Minister of Higher Education of Malaysia, Datuk Mustapha Mohamed, in a press conference in Seremban, stressed that universities should implement a proper control mechanism to curb plagiarism and also implement stringent penalties to plagiarism offenders [2]. He further suggests offenders of plagiarism to be expelled from university. Although the author have yet to find any

The contents of  
the thesis is for  
internal user  
only