Plagiarism Detection Tool for Students' Programming Assignments

Upul Bandara

MS IT 05/10036

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of MSc in Information Technology

March 2011

Declaration

I declare that this dissertation does not incorporate, without acknowledgment, any material previously submitted for a Degree or a Diploma in any University and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organization.

Name of the Student: M.M.C.U.Bandara

Signature of Student: upul Bandara

Date: 14-11-2011

Supervised by

Name of the Supervisor: Dr. GaminiWijayarathna

Signature of Supervisor:

Date:

Dedication

I dedicate this thesis to Dr. Gamini Wijayarathna with a heart full of gratitude. Without his guidance, support, commitment and his uncommendable patience as the supervisor, the successful completion of this work would not have been possible.

Acknowledgments

I would like to extend my sincere gratitude to Dr. Gamini Wijayarathna for his keen supervision, valuable advice, and dearly guidance from the initial stage of the project. It is also my duty to mention here that I have gained a vast knowledge on every aspect of Information Technology by completing this project under the supervision of Dr. Wijayarathna.

Further, I must make this an opportunity to thank my family members, colleagues and friends who helped me to make the project a success in every possible way.

Finally, many thanks go to the academic and non-academic staff members of the University of Moratuwa for their support in various ways.

Abstract

With the rapid development of Internet technology and freely availability of academic materials in electronic format, plagiarism has become a key issue in universities and colleges. Therefore researchers have been developing various tools to address plagiarism.

Plagiarism can be occurred in any academic field, but developing a generic tool to address all kind of plagiarism is not feasible. So we developed a product to detect plagiarism in programming assignments submitted by students those who are taking programming courses in universities and colleges.

Plagiarism detection in programming source codes can be done by various methods. Our proposed solution uses machine learning approach to plagiarism detection. In addition our proposed has been using several learning algorithms and ensemble learning approach to enhanced performance of the system.

At present this system is outperforming some plagiarism detection tools which are based on the machine learning approach to detect plagiarism.

Table of Contents

List of Figures

List of Tables

List of Equations