ABERYSTWYTH UNIVERSITY

PROJECT REPORT

IntelliJ Plugin to Aid With Plagiarism Detection

Author:
Darren S. WHITE
(daw48@aber.ac.uk)

Supervisor:
Chris LOFTUS (cwl@aber.ac.uk)

March 25, 2018

Version: 1.0 (draft)

This report was submitted as partial fulfilment of a BSc degree in

Computer Science (G401)

Department of Computer Science
Aberystwyth University
Aberystwyth
Ceredigion
SY23 3DB
Wales, United Kingdom

Declaration of originality

I confirm that:

- This submission is my own work, except where clearly indicated.
- I understand that there are severe penalties for Unacceptable Academic Practice, which can lead to loss of marks or even the withholding of a degree.
- I have read the regulations on Unacceptable Academic Practice from the University's Academic Quality and Records Office (AQRO) and the relevant sections of the current Student Handbook of the Department of Computer Science.
- In submitting this work I understand and agree to abide by the University's regulations governing these issues.

Name: Darren S. White

Date: March 25, 2018

Consent to share this work

By including my name below, I hereby agree to this dissertation being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Name: Darren S. White

Date: March 25, 2018

Acknowledgements

I am grateful to...

I’d like to thank...

Abstract

Source code plagiarism is an ever-growing issue in academia, primarily in Computer Science. The majority of tools that exist to detect plagiarism only analyse the final piece of code. This paper shows the research and development of a new tool which will track how the code is written. This tool will be an IntelliJ IDEA plugin and will track file changes in the editor. The tracked data will give more of an insight into how plagiarism evolves over the course of development. This method of detection would allow direct identification of the specific pieces of code that were plagiarised.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Overview

Chapter 2

Background

Chapter 3

TODO

Chapter 4

Evaluation

Chapter 5

Conclusions

Appendices

Appendix A

Third-Party Code and Libraries

Appendix B

Code Examples

Annotated Bibliography

[1] R. Lukashenko, V. Graudina, and J. Grundspenkis, "Computer-based plagiarism detection methods and tools," in *Proceedings of the 2007 international conference on Computer systems and technologies - CompSysTech '07*. ACM Press, 2007. [Online]. Available: https://doi.org/10.1145/1330598.1330642

Useful attribute table containing information on tools for detecting plagiarism such as Turnitin, and MOSS.

[2] S. Schleimer, D. S. Wilkerson, and A. Aiken, "Winnowing: local algorithms for document fingerprinting," in *Proceedings of the 2003 ACM SIGMOD international conference on on Management of data - SIGMOD '03*. ACM Press, 2003. [Online]. Available: https://doi.org/10.1145/872757.872770

This went into detail on the ideas behind MOSS, such as document fingerprints and k-grams.

[3] A. Parker and J. Hamblen, "Computer algorithms for plagiarism detection," *IEEE Transactions on Education*, vol. 32, no. 2, pp. 94–99, may 1989. [Online]. Available: https://doi.org/10.1109/13.28038

Details on the plagiarism spectrum. The different levels of plagiarism. Level 0 has no changes and level 6 has the control logic changed.

[4] P. Clough and D. O. I. Studies, "Old and new challenges in automatic plagiarism detection," in *National Plagiarism Advisory Service, 2003; http://ir.shef.ac.uk/cloughie/index.html*, 2003, pp. 391–407.

In-depth details on lexical and structural changes in program code plagiarism.

[5] JetBrains. (2018) Jetbrains/intellij-community: Intellij idea community edition. [Online]. Available: https://github.com/jetBrains/intellij-community

The IntelliJ IDEA Community Edition GitHub repository provides code for plugins. This helped with understanding the IntelliJ Platform Plugin SDK.

[6] U. Bandara and G. Wijayrathna, "Detection of source code plagiarism using machine learning approach," *International Journal of Computer Theory and Engineering*, pp. 674–678, 2012. [Online]. Available: https://doi.org/10.7763/ijcte.2012.v4.555

[7] P. S, R. R, and S. B. B, "A survey on plagiarism detection," *International Journal of Computer Applications*, vol. 86, no. 19, pp. 21–23, jan 2014. [Online]. Available: https://doi.org/10.5120/15104-3428

[8] Flask. (2018) Uploading files - flask documentation (0.12). [Online]. Available: http://flask.pocoo.org/docs/0.12/patterns/fileuploads/

[9] A. J. J. Davis. (2013) Test mongodb failure scenarios with mockupdb. [Online]. Available: https://emptysqua.re/blog/test-mongodb-failures-mockupdb/