Chapter 16
A Source Code Plagiarism Detecting Method Using Sequence Alignment With Abstract Syntax Tree Elements

Hiroshi Kikuchi
The University of Electro-Communications, Japan

Takaaki Goto
The University of Electro-Communications, Japan

Mitsuo Wakatsuki
The University of Electro-Communications, Japan

Tetsuro Nishino
The University of Electro-Communications, Japan

ABSTRACT

Learning to program is an important subject in computer science courses. During programming exercises, plagiarism by copying and pasting can lead to problems for fair evaluation. Some methods of plagiarism detection are currently available, such as sim. However, because sim is easily influenced by changing the identifier or program statement order, it fails to do enough to support plagiarism detection. In this paper, the authors propose a plagiarism detection method which is not influenced by changing the identifier or program statement order. The authors also explain our method's capabilities by comparing it to the sim plagiarism detector. Furthermore, the authors reveal how our method successfully detects the presence of plagiarism.

1. INTRODUCTION

Programming education is one of the most important subjects in computer science courses. In programming exercises, plagiarism caused by copying and pasting leads to problems for fair evaluation. In this study, plagiarism is defined as someone handing in a report or documentation as his or her own original work which was, in fact, written or created by someone else.

Plagiarism is a serious problem. It is difficult for teachers to detect plagiarism with the huge volume of reports they review and, therefore, they are unable to give fair evaluations. Furthermore, students cannot improve their skills if they plagiarize. In addition, teachers must be able to detect plagiarism in order to prevent dishonesty among their students. However, comparing source code or reports manually is very time-consuming.

Teachers would benefit from automatic plagiarism detection, which would allow them more time to focus on their students and to prepare new materials. It is also important to reveal the capabilities of plagiarism detection systems to discourage cheaters. Therefore, techniques for automatic plagiarism detection play an important role in education.

Some methods of plagiarism detection are currently available, such as sim (Gitchell,1999). However, because sim is easily influenced by changing the identifier or program statement order, it fails to provide enough support for plagiarism detection.

In this paper, we propose a plagiarism detection method which is not thwarted by changing the identifier or program statement order. We also explain our method's capabilities by comparing it to sim.
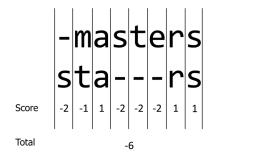
2. BACKGROUND

2.1. Sequence Alignment

Sequence alignment is a method to calculate a correspondence relationship among strings by adding a space or shifting the alphabetic positions. Strings obtained after alignment are also called sequence alignment. The similarity between two strings can be described by a score. In our method, we first obtain tokens by lexical analysis, and then calculate the sequence alignment and compare the obtained similarity scores.

Here is an example. Let s and t be strings. Sometimes the lengths of s and t are different. We insert a gap symbol "-" in order to line up the strings' lengths. This process is called alignment. Figure 1 illustrates an example of calculating the alignment between "masters" and "stars".

Figure 1. Example of a sequence alignment

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/a-source-code-plagiarism-detecting-method-using-sequence-alignment-with-abstract-syntax-tree-elements/222315

Related Content

Action Research Approaches
Karen E. Watkins, Aliki Nicolaidesand Victoria J. Marsick (2018). *Handbook of Research on Innovative Techniques, Trends, and Analysis for Optimized Research Methods (pp. 1-16).*
www.irma-international.org/chapter/action-research-approaches/197725

Investigating the Opportunities to Improve the Thermal Performance of a Case Study Building in London
Yasin Yousefi, Petra Grattonand Dilshad Sarwar (2021). *International Journal of Strategic Engineering (pp. 1-18).*
www.irma-international.org/article/investigating-the-opportunities-to-improve-the-thermal-performance-of-a-case-study-building-in-london/269714

Fostering Professional Dispositions of Preservice Teachers in an Educational Psychology Course
Alpana Bhattacharya (2021). *Handbook of Research on Developing Students' Scholarly Dispositions in Higher Education (pp. 22-41).*
www.irma-international.org/chapter/fostering-professional-dispositions-of-preservice-teachers-in-an-educational-psychology-course/285109

Theory of Constraints and Human Resource Management Applications
Brian J. Galli (2019). *International Journal of Strategic Engineering (pp. 61-77).*
www.irma-international.org/article/theory-of-constraints-and-human-resource-management-applications/219325

The Hook, Woo, and Spin: Academics Creating Relations on Social Media
Megan Jane McPhersonand Narelle Lemon (2017). *Research 2.0 and the Impact of Digital Technologies on Scholarly Inquiry (pp. 167-187).*
www.irma-international.org/chapter/the-hook-woo-and-spin/167442