Acta Cybernetica **19** (2009) 191–216.

Plagiarism Detection in Source Programs Using Structural Similarities

Gergely Lukácsy* and Péter Szeredi*

Abstract

The paper presents a plagiarism detection framework the goal of which is to determine whether two programs are similar to each other, and if so, to what extent.

The issue of plagiarism detection has been considered earlier for written material, such as student essays. For these, text-based algorithms have been published. We argue that in case of program code comparison, structure based techniques may be much more suitable. The main idea is to transform the source code into mathematical objects, use appropriate reduction and comparison methods on these, and interpret the results appropriately.

We have designed a generic program structure comparison framework and implemented it for the Prolog and SML programming languages. We have been using the implementation at BUTE to successfully detect plagiarism in homework assignments for years.

Keywords: plagiarism, program source, graph similarity

1 Introduction and motivation

Comparison of essays and other written materials has been in focus in recent years [27]. Detecting plagiarism in written materials is an issue in education as well as in law procedures. World wide public polls show that two-thirds of university students have used other people's ideas in an impermissible way at least once during their studies. Law disputes include the SCO-IBM debate over the allegedly unauthorised use of portions of the AIX operating system in Linux.

Regrettably, several sites on the Internet provide free or low cost, quick and efficient access to written materials of many types. Unbelievably, sites such as `CheatHouse`[1] or `SchoolSucks`[2] proudly provide tons of essays, dissertations, reports, etc. for students looking for an easy way to have their assignment of some sort fulfilled. We do agree that it is a good idea to get acquainted with the area one

*Budapest University of Technology and Economics (BUTE), Department of Computer Science and Information Theory, 1117 Budapest, Magyar tudósok körútja 2., Hungary, Phone: +36 1 463-2585 Fax: +36 1 463-3157, E-mail: {lukacsy,szeredi}@cs.bme.hu

[1] `http://www.cheathouse.com`

[2] `http://www.schoolsucks.com`