

Chapter 11

iChecker: An Efficient Plagiarism Detection Tool for Learning Management Systems

Samuel P. M. Choi

The Open University of Hong Kong, Hong Kong

Sze Sing Lam

The Open University of Hong Kong, Hong Kong

ABSTRACT

Academic plagiarism is regarded as a serious offense and much effort in the past has been devoted to build stand-alone plagiarism detection systems for a specific language. This paper proposes a new information retrieval-based plagiarism detection algorithm that handles multilingual documents and enables seamless integration with learning management systems. The proposed algorithm employs information retrieval and sequence matching techniques to identify suspected plagiarized sentences and permits parametric control to reduce both false-positive and false-negative results. The full-featured implementation, called iChecker, not only could quickly identify suspected plagiarized works but also ease academics' effort to evaluate the severity of the offence by a quantified measure. Currently iChecker is adopted by over 300 courses (with some having several hundred of students) and has obtained satisfactory results. During 2012 to 2016, iChecker has processed and verified a total of 276,943 documents in English, Traditional Chinese and Simplified Chinese text.

1. INTRODUCTION AND LITERATURE REVIEW

Plagiarized and collusion of assignments and course works has aroused the concern of academics. With a vast amount of information online, the Internet and electronic databases offer tremendous convenience for students to search and download relevant information for completing their assignments and students might be lured to plagiarism. It becomes necessary for academics to put effort to identify the plagiarized works and to properly educate students of intellectual property. However, scanning of students' work for copying are not only time consuming but also impractical sometimes particularly in large classes where assignments are independently marked by multiple tutors.

DOI: 10.4018/978-1-5225-8057-7.ch011

More and more universities introduce the policy of compulsory plagiarism checking of students' works. This motion has a direct impact on the workload of academics. Academics have to be involved in the sanction of students' work and draw a reasonable line between fair use and plagiarism. Except verbatim copying, detection and proof of plagiarism are not trivial. To alleviate the burden of academics, universities can either develop their own plagiarism detection system (PDS) or subscribe services of commercial PDS.


Plagiarism typically involves copying idea of others without permission or appropriately crediting the source (Paredes *et al.* 2007). Martin (1994) has suggested six types of plagiarism, ranging from simple verbatim copying that are easy to detect to completely rewritten ideas of others that are difficult to recognize, as illustrated in Table 1.

It would be extremely difficult, if not impossible, to detect all types of plagiarism mentioned in Table 1. Amongst these six types of plagiarism, word-for-word plagiarism and paraphrasing plagiarism are relatively easy to be reliably detected automatically without human involvement. Except these two types of plagiarism, the others are more elusive and therefore are difficult to develop efficient algorithms for automatic plagiarism detection (Mozgovoy *et al.* 2010; Barrón-Cedeño *et al.* 2013).

Existing plagiarism detection systems are specifically designed for either textual documents or programming codes. Suspected plagiarisms are identified within a local collection of documents (hermetic or collusion detection) or from external sources such as the Internet (Web detection). Both commercial products and open-source freeware are available. Most commercial plagiarism systems (e.g. Turnitin) are proprietary and installed in a server which provides services via the Internet. Open source freeware (e.g. Ferret, Sherlock and WCopyfind) are typically installed on the client-side. Most of the existing textual plagiarism detection systems are designed specifically for English. This paper will focus on server-side, hermetic, textual plagiarism detection for both English and Chinese languages.

Ferret is an open-source plagiarism detection software built by Lyon *et al.* (2001, 2002, 2003) in Computer Science department of University of Hertfordshire. The idea of the algorithm is to evaluate the similarity among the set of three consecutive words (trigrams) in the concerned documents. The

Table 1. Types of plagiarism

<u>Types of plagiarism</u>	<u>Automatic plagiarism detection</u>
Word-for-word plagiarism: direct copying of phrases or passages	
Paraphrasing plagiarism: change of word or syntax	
Plagiarism of secondary sources: quote from secondary sources without look up the original source	
Plagiarism of the form of a source: copy the structure of a source	
Plagiarism of ideas: use the thought from a source	
Plagiarism of authorship: use the works of others	

14 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/ichecker/222310

Related Content

A Mixed Methods Research Study Exploring Customer Satisfaction With the Quality of Service in University Libraries in Sri Lanka

Chaminda Chiran Jayasundara (2022). *Handbook of Research on Mixed Methods Research in Information Science* (pp. 384-418).

www.irma-international.org/chapter/a-mixed-methods-research-study-exploring-customer-satisfaction-with-the-quality-of-service-in-university-libraries-in-sri-lanka/291203

Usage of Social Capital Among Migrant Workers for Their Livelihoods in Malaysia

Bushra Zamanand Rosila Bee Mohd Hussain (2019). *Social Research Methodology and New Techniques in Analysis, Interpretation, and Writing* (pp. 160-189).

www.irma-international.org/chapter/usage-of-social-capital-among-migrant-workers-for-their-livelihoods-in-malaysia/220335

Exploring the Profile and Behavior of Visitors to Crete

Oumayma Mzoughi, George Baltasand George Baourakis (2021). *International Journal of Strategic Engineering* (pp. 55-67).

www.irma-international.org/article/exploring-the-profile-and-behavior-of-visitors-to-crete/269717

Husserlian-Oriented Descriptive Phenomenological Research Methodology

Chau H. P. Nguyenand Howard J. Curzer (2022). *Methodological Innovations in Research and Academic Writing* (pp. 63-81).

www.irma-international.org/chapter/husserlian-oriented-descriptive-phenomenological-research-methodology/291803

Factor Analysis of the Sydney Metro's ITS: System Integration Iteration

Koorosh Gharehbaghi, Kenneth Farnes, Kathryn M. Robsonand Neville Hurst (2021). *International Journal of Strategic Engineering* (pp. 1-13).

www.irma-international.org/article/factor-analysis-of-the-sydney-metros-its/279642