LopezMOSS: An Implementation of Measure of Software Similarity Algorithms

\*Note: Sub-titles are not captured in Xplore and should not be used

Emil John Lopez  
De La Salle UniversityBataan, Philippines  
emil\_lopez@dlsu.edu.ph

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—component, formatting, style, styling, insert (key words)

# Introduction

Plagiarism is an age-old problem in intellectual property law. However, the speed of data transfer and the rise of the Web has made it easier to do the act and has made detection much harder. Furthermore, adjuging someone of commiting the act turns out to be very complicated and can lead the overseer and lawyers in a legal grey area, where it is unsure whether the act violates the law. It is important to examine its definition first to attempt to solve the problem. According to a paper, actions that fall under plagiarism include: (1) turning in someone else’s work, (2) copying someone’s idea without giving credit, (3) not putting quotation marks, and (4) changing words only without changing the structure of the sentence [1].

In educational institutions, software assignments are usually subject to this form of malpractice and so it threatens the integrity of the educational process. In particular, it is subject to the first and second definitions stated above. However, given the number of students each professor must handle, and the large amount of effort required to perform ad hoc comparisons between them, there is a need for a tool that reliably detects plagiarism and can look past obsfucation, reordering, refactoring and other methods of deception [2].

The first significant solution to this problem was introduced in 1994 by an associated professor in UC Berkeley. It used *winnowing,* a local document fingerprinting algorithm that grouped grammatical tokens in groups of some number of tokens, hashed them to minimize storage space, and counted the frequency of each group using their hash value. This is directly derived from other techniques such as Karp-Rabin String Matching [3].

Fingerprinting algorithms calculate numbers (which we call fingerprints) to help identify a document. Conflicts within these numbers usually indicate that some part of a document (or segment of code in this case) is similar to another document. Better fingerprinting algorithms have since been derived from this method, but it remains that the foundation of most of them is either Karp-Rabin String Matching or *n*-gram fingerprinting [4].

This paper will present an implementation of a variation of these key fingerprinting algorithms. In particular, it showcases a version of *n-*gram fingerprinting. The implementation will also have a *graphical user interface* (GUI) that will allow its user to simply pick directories which correspond to projects.

# RELATED WORKS

# METHODOLOGY

## Header

# Using the Template

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**