# DETEKSI KEBERADAAN KALIMAT SAMA SEBAGAI INDIKASI PENJIPLAKAN DENGAN ALGORITMA *HASHING* BERBASIS *N-GRAM*

**aDiana Purwitasari, bPutu Yuwono Kusmawan, cUmi Laili Yuhana**

Lab Semantik *Web*, Teknik Informatika, ITS

Gedung Teknik Informatika, Kampus ITS

Jl. Raya ITS, Kampus ITS, Sukolilo, Surabaya, 60111

E-Mail: a[diana@if.its.ac.id](mailto:diana@if.its.ac.id)

*Abstract*

Abundant cases of plagiarism committed by some intellectual people in the Indonesia’s education fields have became such tragedy. Due to the amount of information which is available online are things that make copy-paste without proper citation cause plagiarism. This paper discusses about how to detect similar sentences which is probable caused by copy-paste. However plagiarism detection still needs further examination such as the existing of citation or not. Winnowing algorithm is used for detecting similar sentences between text files which is treated as a common subsequence problem. The algorithm finds document fingerprinting by changing sequence of N-grams from text into a set of hash values. If copy-paste sentences are found then both of text files must have the same document fingerprinting. Experiment has been done to observe the capability of detecting similar sentences by analyzing on value variations of n-gram, prime base b for hashing, window w, and threshold for determining plagiarism indication.

**Keywords:** common subsequence problem, winnowing algorithm, document fingerprinting, n-gram, hashing.