A SIMILARITY MEASURE ANALYSIS BASED IMPROVED APPROACH FOR PLAGIARISM DETECTION

MAULI JOSHI[1] & KAVITA KHANNA[2]

[1]Research Scholar, Department of CSE, PDM College of Engineering for Women, Bahadurgarh, Haryana, India

[2]Associate Professor, Department of CSE, PDM College of Engineering for Women, Bahadurgarh, Haryana, India

ABSTRACT

Web Mining is one most useful concept used by the each user knowingly or unknowingly. As user searches some text over the search engine, the web information retrieval is performed. When a user writes some document, research paper, report, most of the content part is taken from the web. In such case, it is required to analyze the effective working of a user. The concept of detection of user contents over the web is called plagiarism. Plagiarism is a kind of content based or the logic based theft. The presented work is about to detect these plagiarism contents over the web. To perform this detection we have defined a statistical approach. At the initial stage, the content filtration is performed to reduce the query size. After that the similarity measures are been used to perform the content match over the web. The obtained results show the effective detection of user contents over the web.

KEYWORDS: Web Mining, Statistical Measure, Plagiarism, Query Filtration